

GeoMetaMatica

Inleidingen van de studiedag over metagegevens voor geografische informatie

Subcommissie Geo-Informatie Modellen, Utrecht, 18 mei 2004

L. Heres (redactie)

NCG Nederlandse Commissie voor Geodesie

Delft, mei 2004

GeoMetaMatica
L. Heres (redactie)
ISBN 90 6132 285 5

Vormgeving en productie: Bureau Nederlandse Commissie voor Geodesie
Druk: Optima Grafische Communicatie, Rotterdam

Bureau van de Nederlandse Commissie voor Geodesie
Bezoekadres: Kluyverweg 1, 2629 HS Delft
Postadres: Postbus 5058, 2600 GB Delft
Tel.: 015-2782819
Fax: 015-2781775
E-mail: ncg@lr.tudelft.nl
Website: www.ncg.knaw.nl

De Nederlandse Commissie voor Geodesie (NCG) is een onderdeel van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

Inhoudsopgave

<i>Voorwoord</i> Ir. Luc Heres	v
<i>Ontwikkelingen voor metadata concepten</i> Prof.ir. Henri J.G.L. Aalders	1
<i>GeoKey - de sleutel tot uw 'Map Services'</i> Ing. Ronald C. Bottelier	14
<i>Sleutelpositie voor meta-informatie en meta-informatiebeleid in de Geo-Informatie Infrastructuur van Rijkswaterstaat</i> Dr. Michel J.M. Grothe en ir. Paul van Asperen	20
<i>Informatiefabrieken en meta-informatie</i> Ir. Hein J.M. Corstens	33
<i>SuperGlue: De lijm tussen alle metadata</i> Bert Oosterhof	40
<i>Metadata en gedistribueerd GIS</i> Drs. Bert Vermeij	44
<i>De Universal Data Store en de rol van meta-informatie</i> Ir. Remco J.J.H. van Eijndt	50

Voorwoord

Ir. Luc Heres
Rijkswaterstaat, Adviesdienst Verkeer en Vervoer, Heerlen
l.heres@avv.rws.minvenw.nl

Meta-informatie, iets nieuws onder de zon?

De term meta-informatie is relatief nieuw. Bernard Plagman, auteur van Data Dictionary and Directory Systems heeft de term als eerste gebruikt in een artikel uit 1971, zegt men. En hoewel je altijd wat voorzichtig moet zijn met dit soort claims, is het duidelijk dat de term uit de informatietechnologie en met name uit de databasehoek afkomstig is.

Bestond er voor die tijd dan geen meta-informatie? Wel zeker, alleen noemde men het niet zo. Kaartenmakers bijvoorbeeld, spreken vanouds over 'randinformatie'. Daarmee bedoelen ze de gegevens die op de kaartrand staan gedrukt, buiten het eigenlijke kaartbeeld zelf, zoals titel, uitgever, schaal en legenda.

Ook is meta-informatie niet specifiek voor geografische informatie, maar speelt het overal en altijd een rol: daar waar informatie is, daar is meta-informatie. Meta-informatie beschrijft de *context* waarin bepaalde informatie gelezen moet worden. Het is daarom een *relatief* begrip: als ik zeg dat iemands telefoonnummer 563426 is, dan is de term 'telefoonnummer' meta-informatie is ten opzichte van '563426' omdat het de context beschrijft waarin dat getal gelezen moet worden.

Business meta-informatie versus technische meta-informatie

David Marco, auteur van een aantal boeken op het gebied van Meta Data Management [1], onderscheidt een tweetal soorten metagegevens: *business metadata* en *technische metadata*. Onder business metadata verstaat hij onder andere de gegevens die men nodig heeft als men een bestand wil beschrijven of ernaar wil *verwijzen* (in een catalogus bijvoorbeeld). Het gaat bijvoorbeeld om gegevens als titel, auteur, jaar van uitgave, schaal en dergelijke.

Technische metagegevens zijn data die men nodig heeft als men iets met de informatie wil *doen*, interpreteren of combineren met andere gegevens bijvoorbeeld. De klassieke kaartlegenda valt in deze categorie: een legenda beschrijft de semantische relatie tussen de begrippen uit de 'buitenwereld' en de grafische elementen uit het kaartbeeld (de 'binnenwereld').

Ook digitale bestanden hebben hun 'legenda', alleen heet die meestal niet zo. Vaak spreekt men over het 'data schema', of over de 'repository' (schatkamer). Ook de repository legt een verband tussen de 'buitenwereld' en de 'binnenwereld'. De binnenwereld bestaat in dit geval uit datatypes, gegevensvelden, waardedomeinen en datastructuren. Een wereld die men moet kennen als men deze wil combineren met de informatie uit een ander bestand met zijn eigen datatypes en -structuren.

Studiedag GeoMetaMatica

De afgelopen jaren wordt het begrip 'meta-informatie' ook binnen de geo-informatie gemeenschap steeds vaker genoemd. Er zijn inmiddels Europese en internationale normen beschikbaar. Het leek de Subcommissie Geo-Informatie Modellen van de Nederlandse Commissie van Geodesie een goed idee een studiedag te organiseren over dit onderwerp waarbij zowel de huidige toepassingen als de toekomstige ontwikkelingen de revue zouden passeren.

De voordrachten van deze studiedag zijn in deze publicatie gebundeld. De studiedag is gehouden op 18 mei 2004 in het Mobilion in Utrecht.

De geschiedenis van meta-informatie

Henri Aalders van het Onderzoeksinstituut OTB van de Technische Universiteit Delft belicht de historie van meta-informatie en legt uit hoe de momenteel gebruikte begrippen én termen tot stand zijn gekomen. Vervolgens gaat hij in op de theorie van metagegevens en laat zien dat ook de oude Grieken zich in feite al met meta-informatie bezig hielden, ook al noemden ze dat niet zo. Daarna brengt hij de internationale ontwikkelingen ter sprake, met name op het gebied van de stroomlijning en standaardisatie. Daarbij laat hij zien dat deze activiteiten niet alleen staan maar passen in grote metadata initiatieven zoals de Dublin Core Metadata Initiative en de Learning Object Metadata Working Group. Hij schetst hoe de CEN-activiteiten zich verhouden tot die van ISO. Tenslotte maakt hij een rondje langs alle continenten en beschrijft daar de huidige stand van zaken.

De sleutel tot geo-informatie

De firma Geodan timmert al een aantal jaren aan de weg met GeoKey, een hulpmiddel om geografische *business meta-informatie* op een efficiënt manier in te vullen en te beheren. Aanvankelijk was GeoKey gebaseerd op specificaties die intern bij Rijkswaterstaat werden gehanteerd. Bij het verschijnen van de CEN-norm zijn deze specificaties bijgesteld. Ronald Bottelier van Geodan gaat in op het GeoKey product en schetst de nieuwe mogelijkheden die de web-services architectuur en standaarden van het OpenGIS Consortium bieden voor een toekomstige GeoKey versie, gebaseerd op open standaarden.

Meta-informatie beleid en geo-informatie infrastructuur

Het hebben van goede technologische hulpmiddelen mag belangrijk zijn, als een organisatie er niet op een uniforme manier mee aan de slag gaat, leidt het tot niets. Om dat te voorkomen is een geïntegreerd geheel aan technische, organisatorische én beleidsmatige middelen nodig: een meta-informatie Infrastructuur

Michel Grothe van de Adviesdienst Geo-informatie en ICT (voorheen de Meetkundige Dienst) gaat in op de 'Meta-Informatie Infrastructuur' van Rijkswaterstaat. Richtinggevend voor de inrichting van deze infrastructuur is het in 2002 geformuleerde meta-informatiebeleid. Allereerst omschrijft hij de meta-informatie infrastructuur binnen Rijkswaterstaat en introduceert het meta-informatiebeleid. Daarna gaat hij in meer detail in op de onderdelen en invulling van de meta-informatie infrastructuur voor de geo-informatievoorziening van Rijkswaterstaat, te weten beleid en afspraken, standaarden, gegevens, systemen en technologie en de organisatie.

Deze Meta-informatie Infrastructuur (meestal afgekort als MII of MI²) richt zich momenteel voornamelijk op *business meta-informatie*. Maar het lijkt geen twijfel dat de *technische meta-informatie* in een volgende fase aan de orde komt.

Informatiefabrieken en de integratie van data

Hein Corstens is directeur en oprichter van Urbidata, een informatiseringbedrijf dat gericht is op de integratie van de informatievoorziening bij bedrijven en overheidsorganisaties. Hij ziet de 'informatiefabriek' als de logische industriële opvolger van de huidige vaak nog ambachtelijke

manier van informatieverwerken. Het verschil tussen een informatiefabriek en een ambachtelijk proces is dat de bewerkingen veel verder geformaliseerd zijn. Een centrale rol in zo'n fabriek is de zogeheten *Universal Data Store*, een gegevensmagazijn waar de data op een samenhangende manier worden opgeslagen. Een onlosmakelijk onderdeel daaruit is de *Meta-Database*: een catalogus met informatie over de gegevens in de bronsystemen plus alle integratie- en transformatieprocessen. Hiermee worden zowel het gegevensbeheer als de informatievoorziening ondersteund.

De Universal Data Store en de rol van meta-informatie

Remco van Eijndt gaat wat dieper in op de rol van metadata in de Universal Data Store:

- Wat zijn metadata? Metadata leggen de structuur en semantiek vast van een dataset. Op diverse niveaus kunnen deze 'gegevens over gegevens' worden vastgelegd. Metadata is dan ook een relatief begrip, want ook over metagegevens zelf en over metametagegevens kunnen gegevens worden vastgelegd.
- Metadata bij transformaties. Bij de traditionele transformatie van (ruimtelijke) gegevens door middel van 'structural transformation' treedt gegevensverlies op. Dit kan voorkomen worden door gebruik te maken van 'semantic translation'.
- Metadata in de UDS: een catalogus met informatie over de gegevens in de bronsystemen en het systeem zelf, alsmede alle integratie- en transformatieprocessen.

Superglue, de lijm tussen alle metadata bronnen

Bert Oosterhof is Technisch Directeur Europa van Informatica, een Amerikaanse softwareleverancier, die actief is op het terrein van datawarehousing, business-intelligence en metadata-management. Informatica is onder andere producent en leverancier van het data-integratie platform PowerCenter.

Sinds kort levert Informatica ook SuperGlue. Hiermee kan men metadata vanuit diverse applicaties integreren, samenvoegen, relateren en beheren. Het is enerzijds een Metadata Warehouse, waarbij bijv. analyses en rapportages gemaakt kunnen worden, en anderzijds een open en uitbreidbaar platform voor metadata-management. In zijn bijdrage gaat Bert Oosterhof tevens in op de vraag hoe producten als PowerCenter en Superglue, die ontwikkeld zijn voor generieke informatiesystemen, succesvol kunnen worden ingezet bij het bewerken en beschrijven van geografische meta-informatie.

Metadata en gedistribueerd GIS

Webservices technologie maakt echte gedistribueerde GIS systemen mogelijk. Data blijft beheerd bij de bron, via internet tappen gebruikers van GIS applicaties de benodigde gegevens af. Metadata is hiervoor onmisbaar. De recente versies van de GIS software van ESRI sluiten volledig aan op gedistribueerde systeemconcepten.

ArcCatalog is een onderdeel van ArcGIS en bedoeld om geografische metagegevens te kunnen beschrijven en beheren. Veel metagegevens worden direct uit het bestand zelf overgenomen. ArcCatalog richtte zich aanvankelijk voornamelijk op de *business meta-informatie*, maar in de latere versies wordt de *technische meta-informatie* steeds belangrijker. De allernieuwste versie richt zich bijvoorbeeld ook op de vastlegging van de ontologie van een bestand en de beschrijving van bewerkingsprocessen en informatiestromen (dataflows) tussen bestanden onderling.

Bert Vermeij van ESRI-Nederland belicht de laatste ontwikkelingen rond metadata in ESRI software, met name richting geoservices.

Referentie

- [1] David Marco; Building and Managing the Meta Data Repository: A Full Lifecycle Guide; John Wiley & Sons; Book en CD-ROM editie (Juli 17, 2000); ISBN: 0471355232

Ontwikkelingen voor metadata concepten

Prof.ir. Henri J.G.L. Aalders
Technische Universiteit Delft, OTB
h.j.g.l.aalders@geo.tudelft.nl
Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen

Introductie

Oorspronkelijk werd geografische informatie geproduceerd en gebruikt in een geo-omgeving met deskundigen die zich hadden gespecialiseerd in de geografie, kartografie, geodesie, fotogrammetrie, remote sensing, zeebodemkartering, hydrografie, geologie, bodemkunde, ruimtelijke planning, architectuur, civiele techniek, milieukunde, etc. Hun werk resulteerde in de kartografische representatie van de werkelijkheid eerst in kaartvorm en later als een digitaal computer model. Combinaties van verschillende ruimtelijke datasets in verschillende schalen, projecties en coördinaatsystemen, inhoud en verschijningsvorm creëert een gezamenlijk computermodel van de aarde als geheel of in delen. Dit omvat de aggregatie van digitale geografische, topografische en thematische gegevens (zoals ortho-fotokaarten, stadskarten, lucht- en ruimtelijke fotobeelden, (3D-)terrein modellen met gravimetrische, fotogrammetrische en kartografische informatie) geïntegreerd en geordend met een gemeenschappelijk concept zoals een geodetische basis, inhoud, ontwerp en referentie informatie en opgeslagen in digitale vorm.

Om geografische informatie geschikt te maken voor gebruikers, pasten deze deskundigen maatregelen toe met een conceptuele en methodologische basis, wettelijke regelingen en normen voor formattering en gegevens overdracht. Ook zijn implementaties op het gebied van hard- en software ontwikkeld en technologieën om de gegevens over te dragen naar gebruikers (Moellering, 1991), (Moellering, 1994), (Moellering 1997), voor de opslag, analyse en verwerking en kartografische presentatie van gegevens.

In het dagelijkse leven zijn we gewend om de labels van gekochte producten te lezen om verschillende soorten informatie te verkrijgen, zoals het land van herkomst, het toegepaste productieproces, de laatste geldigheidsdatum, garantiebepalingen, toegestane gebruik, enz. Deze informatie is gebonden aan wettelijke bepalingen. Voor geografische informatie bestaan dergelijke bepalingen niet. Maar nu we in een digitale geo-informatie infrastructuur werken, die deze informatie beschikbaar stelt aan lekengebruikers, is universele informatie nodig over de locatie, de evaluatie, het verkrijgen en het toepassen van de gegevens.

Beschrijvingen van geografische datasets bestaan nu al enige tijd. Meestal zijn deze beschrijvingen ontstaan in verschillende omgevingen, in verschillende landen en voor verschillende doeleinden, waardoor ze vaak niet met elkaar zijn te vergelijken en weinig overeenkomst vertonen en niet uitwisselbaar zijn; tenminste in een internationale omgeving. Dit is de aanleiding geweest om een internationale norm te ontwikkelen voor ruimtelijke gegevenbeschrijvingen: ISO19115:2003, Geographic Information – Metadata.

Doelstelling van de ontwikkeling van ISO19115: 2003, Geographic Information - Metadata was o.m.:

- ondersteunen van het gebruik van ruimtelijke informatie;
- samenwerken met de informatie technologie bij de ontwikkeling van de norm;
- wereldwijde toepassing van ruimtelijke gegevens bevorderen over de landsgrenzen heen en in een meertalige omgeving;
- baseren op de fundamenteën van nationale en regionale ontwikkelingen en ervaringen door een diepgaande analyse uit te voeren en implementaties te testen.

ISO19115:2003 Metadata definieert een set metadata elementen en karakteristieken, evenals het schema om metadata volledig te documenteren. De norm kan op alle ruimtelijke gegevens worden toegepast: datasetseries, datasets, individuele ruimtelijke objecten, en attributen. De norm definieert ook een minimum set metadata, maar laat ook optionele metadata elementen toe om een meer uitgebreidere dataset te kunnen beschrijven.

In de IT wereld bestaan nog twee andere prominente metadata initiatieven:

- Dublin Core Metadata Initiative (DCMI, Dublin Ohio, U.S.A, metadata element set (zie <http://www.xml.com/pub/2000/10/25/dublincore/>) is bedoeld om zoekopdrachten naar verschillende typen onderwerpen uit te voeren en te verkrijgen (ofwel de basis: in elkaar grijpende bouwstenen). Het kan gezien worden als een voorbeeld metadata zoektaal dat de gebruiker laat navigeren door van elkaar afwijkende onderwerpen, in verschillende talen en culturen. Het wordt veel toegepast in gouvernementele omgevingen, musea, bibliotheken, milieuwetenschappen, drukkerijen, handel, gezondheidszorg, financiële sector, enz., waaruit het succes wel blijkt.
- Institute for Electrical and Electronic Engineers (IEEE) Learning Object Metadata (LOM) Working Group, zie <http://grouper.ieee.org/p1484/doc/wg12/LOM-WD3.htm> voor technische toepassingen.

Geschiedenis van metadata

Voor velen zijn de concepten gegevensbeschrijving, metadata, legenda en marginalia hetzelfde: allen beschrijven ze de eigenschappen en karakteristieken van datasets.

Vroeg in de jaartelling pasten de Griekse geografen al beschrijvingen van kaartobjecten toe. Veel middeleeuwse geografen maakte prachtige tekeningen om hun visie van de wereld in kaarten weer te geven: veelal met religieuze motieven en altijd met de auteursnaam. Vanaf het midden van de 18^e eeuw topografische, zee- en navigatiekaarten lieten beschrijvingen en verklaringen zien om een betere kaartinterpretatie mogelijk te maken; dit werd kaartmarginalia genoemd, omdat de informatie vaak in de marge van de kaart werd vermeld. Marginalia vermeldde o.m. verkenningsdata, begrenzingen, coördinaatsystemen, schaalinformatie, kwaliteit, uitgever of eigenaar, enz.

ISO/TC 211, de auteur van de ISO19115:2003, Metadata norm, omschrijft metadata met: "Gegevens over de inhoud, kwaliteit, conditie en andere karakteristieke gegeven van een dataset".

De ICA Commissie voor Standaarden voor Ruimtelijke gegevens gebruikt de omschrijving: "informatie over een ruimtelijke dataset voor een beter begrip van ruimtelijke gegevens".

Volgens deze definitie bevat metadata: templates voor overdrachtformaat, tellingen van data items in de overdracht, conceptuele gegevensmodellen, catalogi, referentiesystemen, kwaliteitsrapportage en de logische beschrijving van de metafile.

De term **metadata** * – of ook gebruikt in andere vormen, zoals *meta data* (McIntosh, 1968), *meta-data* (Homer, 1978), (Ziegler, 1978), (Weber, 1978), (Schelling, 1978) en *metadata* (DLC, 1979) – dook op in de Computer wetenschappen/Informatie literatuur in 1968. Sindsdien wordt de term gebruikt in alle drie vormen: metadata werd populair in Europa in het begin van de zeventiger jaren en werd later in Amerika gebruikt in militaire documenten als één woord: metadata.

Theoretische aspecten voor metadata

Volgens de concepten uit de verzamelingentheorie, kan een verzameling geografische gegevens verschillende andere homogene verzamelingen bevatten met een of meerdere exemplaren van objecten, attributen en/of relaties. Daarmee kan men onderscheid maken in verschillende hiërarchische niveaus van verzamelingen. Bovendien blijkt het voor de toepassing van dergelijke verzamelingen noodzakelijk om ook de beschrijving van de inhoud van iedere verzameling te geven. Naast een gegevensbeschrijving van de verzameling bevat de metadata ook een beschrijving van de beheerorganisatie, en het gegevensonderhoud, van zowel de verzameling van geografische gegevens, als van de metadata van de verzameling. Complete metadata beschrijvingen van de structuur, inhoud en nauwkeurigheid van alle gegevensverzamelingen is een belangrijke eis voor het ontwerpen van een gegevensbestand om bescherming en misbruik van de verzameling te voorkomen.

Een metadatamodel omvat de definitie van verplichte informatie, mogelijke metadata en conditionele metadata (metadata die verplicht is onder bepaalde omstandigheden). Bovendien moet de minimale metadata gegevensset worden weergegeven (meestal een opsomming van de verplichte metadata).

Metadata kunnen worden opgeslagen vanuit verschillende principes, bijvoorbeeld vanuit:

- de managementstrategie;
- financieel oogpunt;
- institutionele of technische achtergrond.

De beste tijd om metadata te definiëren en te verzamelen is tijdens het verzamelen van de gegevens zelf, omdat afwachten tot na de gegevens invoer zal resulteren in een slechtere beschrijving van de te verzamelen gegevens of tot een kostenverhoging omdat later aanvullende informatie moet worden verzameld. Ook kan het zijn, dat achteraf bepaalde details worden vergeten.

Het toepassen van gegevens uit verschillende bronnen, en mogelijk ook het toepassen van processen op gegevens op verschillende plaatsen door een GRID-computing benadering met gedistribueerde verwerking maakt nieuwe en innovatieve applicaties mogelijk. Dit zou kunnen uitmonden in een Virtueel GIS (V-GIS omgeving), waarin de Digitale Wereld, opgeslagen in bestanden, naadloos kan worden bestudeerd, gebruikmakend van het gehele, gecombineerde spectrum van karakteristieken, die opgeslagen zijn in de afzonderlijke bestanden en door processen die op verschillende computers beschikbaar zijn. Dit concept is onderwerp van discussie ge-

* De Metadata Company registreerde de term 'Metadata' in 1986. De Metadata Company verkoopt goederen en diensten gekoppeld aan metadata. Omdat de FGDC geen goederen of diensten verkoopt, vindt ze dat ze de registratie niet overtreedt zoals bepaald door het Department of the Interior, Office van de Solicitor (zie FGDC Official statement op http://www.fgdc.gov/metadata/meta_trademark.htm). Omdat de registratie goederen en diensten betreft, hebben verschillende bedrijven die metadata producten vervaardigen de neiging om de term metadata te vermijden en neigen ertoe de term metadata te gebruiken of zelf andere termen zoals gegevenseigendom of gegevenscatalogisering.

weest tijdens de G8 conferentie in 2001 met de Verenigde Naties, verschillende staatshoofden, non-profit organisaties en internationale, particuliere instellingen.

Hiërarchie in standaardisatie

In sommige landen wordt de toepassing van normen en standaarden bij wet geregeld (bijvoorbeeld Portugal en Rusland), terwijl in andere landen het gebruik ervan wordt overgelaten aan het particuliere initiatief (bijvoorbeeld Nederland en Zweden). Landen waar de toepassing bij wet wordt geregeld maken onderscheid tussen:

- internationale normen, meestal ontwikkeld door de ISO, terwijl regionale normen worden ontwikkeld door een regionale internationale organisatie zoals Comité Européen de Normalisation (CEN). Veel nationale regeringen en norminstituten nemen deze normen over in hun nationale normen bibliotheek.
- nationale normen aangenomen door de (Staats-)Normcommissie, een normalisatiewet of een nationaal decreet (Koninklijk Besluit);
- domeinstandaarden die worden gebruikt door verschillende (overheids-)organisaties of een groep organisaties in hetzelfde werkgebied;
- toepassingsstandaarden, toegepast binnen een specifiek bedrijf.

Veel normen staan niet op zichzelf maar maken onderdeel uit van een serie normen die ieder hun eigen aspect van de normalisatie in een bepaald toepassingsveld beschrijven (ook wel een normen complex of serie normen genoemd, zoals de ISO normen 191xx voor Geografische informatie).

Omdat normen onder andere ook de objectdefinities bevatten (in de geografische informatie ook wel ontologie genoemd) en het moeilijk is om voor alle voorkomende toepassingen de definities te harmoniseren (of identiek te maken), wordt vaak uitgegaan van een hiërarchie in de objectdefinities, waarbij bij een lager niveau in de hiërarchie de genormaliseerde definitie uit een hoger niveau wordt geïmplementeerd en eventueel nader gepreciseerd. Daarbij wordt voorgaande indeling aangehouden. Overdracht van gegevens kan dan ook alleen maar plaatsvinden binnen een bepaald niveau of hoger.

Normen kunnen verschillende aspecten voorschrijven voor de levering van gegevens, ontologie, kwaliteit, conceptuele modellen, enz. Normen worden gebaseerd op gegeneraliseerde resultaten van wetenschappelijk, technisch en toegepast onderzoek en als resultaat van de noodzaak in de praktijk. Bovendien is de intentie dat er publiek voordeel mee gedaan kan worden. Normen voor metadata vormen daar geen uitzondering op.

Metadata modulariteit en flexibiliteit

Een ander concept bij metadata normen is de **modulariteit** (Duval, 2002): metadata modulariteit is een fundamenteel principe dat ontwerpers toestaat nieuwe samenstellingen te creëren, gebaseerd op bestaande metadata schema's: in een modulair systeem kunnen metadata elementen uit verschillende gebieden worden samengevoegd zowel syntactisch als semantisch. Daarmee kunnen metadata gegevens hergebruikt worden in grotere verbanden zelfs als de semantiek niet overduidelijk te combineren is, maar er wel een gemeenschappelijk syntactische fundament onder ligt.

Metadata modulariteit kan er ook voor zorgen in een specifieke noodzaak binnen een bepaalde toepassing wordt voorzien. Zo zal bijvoorbeeld de *identificatie* bij elke dataset voorkomen, terwijl de *graad van bewolking* alleen bij luchtopnames en remote sensing beelden wordt weergegeven.

Bovendien verlangt metadata modulariteit voldoende verfijning om meer specifieke informatie kenbaar te maken; bijvoorbeeld *auteur, illustrator, componist of beeldend kunstenaar* zijn een verbijzondering van een meer algemene term *ontwerper*.

Toepassing van meerdere talen en culturen wordt ook steeds belangrijker nu Internet zich aandient voor de distributie van (ruimtelijke) gegevens: Internet zal aan betekenis inboeten indien we niet in staat zijn om gebruikers in hun moedertaal te bedienen. Hier komen twee tegengestelde begrippen als **internationalisering** en **lokalisering** naar voren. Zoeken naar gegevens via Internet over de gehele wereld is het best gediend bij een internationalisering (met gebruik van gemeenschappelijke afspraken, taal en karaktersets), terwijl de behoefte van een specifieke gemeenschap beter is gediend met lokale afspraken.

Een oplossing kan zijn om in de internationale metadata architectuur automatische vertalingen te laten plaatsvinden in verschillende omgevingen, culturen en talen. DCMI (Dublin Core Metadata Initiative) onderhoudt hiertoe een lijst met vertaalde documenten en de European Workshop on Learning technologies doet hetzelfde voor de LOM specificaties.

Tenslotte vormen namespaces een integraal onderdeel van de Web infrastructuur. Namespaces worden gebruikt in bevragingen via het Internet als onderdeel van een modulaair metadata-systeem. Zo kan bijvoorbeeld in een bevraging via het Web de URL de prefix bevatten *dc* om een Dublin Core metadata gestructureerde elementset te zoeken terwijl *lom* een zoekopdracht van metadata volgens de IEEE-LOM structuur uitvoert.

Nominale grondslag

Iedereen neemt zijn omgeving waar en creëert daar een model van volgens eigen inzichten. Ruimtelijke bestanden streven een objectieve digitale voorstelling van de werkelijkheid na door IT-oplossingen toe te passen, maar ook ruimtelijke computer analyses met grafische voorstellingen. Echter, iedere beschrijving van de werkelijkheid is altijd een abstractie van de objecten die waarneembaar zijn in de werkelijkheid en die abstractie is altijd subjectief: het is een van de vele voorstellingen die we ervan kunnen maken. Bovendien is het niet een exacte kopie van de werkelijkheid: zaken worden benaderd en gedeeltelijk voorgesteld, gesimplificeerd of soms zelf niet voorgesteld. Om er zeker van te zijn dat een gegevensbestand niet misbruikt wordt door verkeerde interpretatie, aannames of beperkingen moet elk bestand volledig worden beschreven voor wat de inhoud betreft: de metadataset.

De nominale grondslag wordt in het Engels: 'UoD, Universe of Discourse' genoemd. In de Verenigde Staten van Amerika, wordt meestal de term 'abstract view of the universe' gebruikt; dat geeft echter niet het selectieproces (zie hieronder) weer en dat is nu juist een integraal onderdeel van de nominale grondslag definitie. ISO/TR 9007, 1987 (ISO, 1946) definieert de UoD als "alle objecten of fenomenen van interesse voor een bepaalde toepassing, die de werkelijkheid in een abstracte, karakteristieke vorm beschrijven door de thematische, geometrische en temporele semantiek". Daarvoor wordt een **ontologie** gebruikt, waaronder in de geografische informatie wordt verstaan: *een definitie, classificatie en structuur van objecten, als een limitatieve opsomming van eenduidige uitgekozen - beschouwd vanuit een specifieke toepassing - concepten* (Uitermark, 2001). Deze definitie volgt de doctrine van de Griekse filosoof Parmenides (500 - 450 BC) en de Indiase Nyāya-Vaiśeṣika filosofie school op het gebied van ontologie. Op deze wijze bevat de nominale grondslag een limitatieve set karakteristieken van geselecteerde objecten uit de werkelijkheid met hun beschrijvingen. Objecten in de werkelijkheid worden in een gegevensbestand gerepresenteerd en daardoor bestaat er een één-op-één relatie tussen de bestand objecten en de werkelijkheid. Het proces om de objecten uit de werkelijkheid in een bestand te representeren bestaat uit drie stappen:

1. selectie van objecten uit de werkelijkheid die in aanmerking komen voor een bestandrepresentatie. Door deze selectie, uitgevoerd door de potentiële gebruiker, wordt het bestand direct geënt op de gebruikers wensen en waarschijnlijk allen voor zijn – beperkte – doeleinde toepasbaar. Dat geeft een productbeeld, zowel emotioneel als cognitief.
2. abstractie van die objecten, dat resulteert in de inhoud van het te vormen bestand;
3. meting van de objecten en opname in het bestand volgens een gespecificeerde werkwijze.

Toepassing van metadata

Geen enkele metadataset kan alle functionele eisen voor alle toepassingen behappen, terwijl met de ingebruikname van het Web de toegang tot datasets steeds eenvoudiger wordt. Daarom wordt het steeds belangrijker om modellen te ontwikkelen voor bestandbeschrijvingen en bijbehorende zoekstrategieën, door het ontwikkelen van:

- *Conceptuele modellen* die de nominale grondslag definiëren, nodig voor de representatie van de syntax en semantiek van metadata, als vervanging van de tekstuele beschrijvingen, lijsten met objecten en attributen, die momenteel nog zo vaak in metadata worden aangetroffen en gemakkelijk tot onjuiste interpretatie leiden. Een nominale grondslag bevat daarom object(-typen) met hun attribuut(-typen), en de relatie(-typen) tussen die objecten met hun attribuut(-typen) en de betreffende domeinen voor alle attributen evenals het gedrag van de objecten in een objectgeoriënteerde model.
- *Syntax en Semantiek*. Syntax gaat over de vorm, semantiek over de betekenis. Er moeten over beiden afspraken worden gemaakt om metadata te kunnen overdragen. Syntax definieert de wijze waarop de gegevens worden overgedragen, terwijl semantiek definieert wat er wordt overgedragen. Er bestaat nog een derde component in de interoperabiliteit, dat hierboven uitstijgt: de inhoudelijke woordenschat (lexicon). Die kan zijn in natuurlijke taal, zoals in het Nederlands of Engels, maar ook in specifieke formele taal die de betekenissen van beschrijvingen in de overdracht verder beperkt en daarmee de kwaliteit van de overdracht verhoogt.
- *Toepassingsprofielen*, een combinatie van metadata elementen uit verschillende schema's die de principes van modulariteit en uitbreidbaarheid mogelijk maakt, maar is toegespitst op een specifieke toepassing. Dit wordt mogelijk door:
 - definitie van de *cardinaliteit*, waarbij de beperkingen elk metadata element worden aangeduid (als verplicht, facultatief of conditioneel);
 - gebruik van *waarde beperkingen* voor specifieke toepassingen;
 - *relatie- en afhankelijkheidspecificaties* tussen verschillende object(-type)en.
- *Declaratie van namespaces* als er meerdere worden ondersteund.

Een van de eerste stappen bij de ontwikkeling van een gegevensmodel is de analyse van de gebruikerswensen, gebaseerd op het toekomstige potentieel gebruik van de gegevens, resulterend in een set externe schema's voor iedere mogelijke toepassing. Het conceptuele schema voor het informatiesysteem is de combinatie van al deze externe schema's. Echter, informatiesystemen en hun toepassingen hebben een beperkte levensduur en waarschijnlijk zullen nieuwe ideeën en technologieën vereisen dat nieuwe conceptuele modellen worden ingevoerd.

Metadata elementen set

Metadata normen zijn een middel om de inhoud van informatiesystemen te beschrijven op verschillend detailniveau. Het meest vage niveau is een *samenvatting of overzicht* met een korte inhoudsopgave van het informatiesysteem, meestal in natuurlijke taal en ongestructureerd. Dit

geeft de leverancier van de metadata wel veel vrijheid, maar het gebruik wordt dan wel afhankelijk van de mogelijkheden van de gebruiker, omdat in de natuurlijke taal inconsistente en dubbelzinnige uitdrukkingen voorkomen, die vaagheid en onbegrip in de hand werken ondanks een goed gedefinieerde grammatica en daardoor automatische interpretatie bemoeilijken. Een lijst met object- en attribuutbenamingen inclusief de relatietypen en de bestaande hiërarchieën geeft al een veel betere automatische interpretatie; vooral als ook een conceptueel schema wordt meegeleverd. Daarom worden in de metadata elementenset ontwikkeld; men kan daarin drie typen element onderscheiden:

1. *Metadata van metadata*, die gaan over de beschrijving en definitie van metadata (let wel dit zijn niet de metadata zelf, maar gegevens om de metadata te begrijpen), zoals de gebruikte taal voor de metadata, de karakterset en het coderingsysteem (bijvoorbeeld ASCII), de referentiesystemen voor temporele, thematische en metrische attributen, het referentiekader voor de kwaliteitsdefinities en de administratieve informatie voor de metadataset (contactpunt, bijhouding enz.).
2. *Directory metadata* met de identificatie van de dataset, de afkomst informatie van de dataset, de plaats waar de dataset is te vinden en onder welke voorwaarden de dataset is te verkrijgen, adres van eigenaar, beheerder en distributeur en hun rol t.a.v de dataset, contactpunt voor informatie over de dataset, de gebiedsbeschrijving: zowel ruimtelijke als temporeel en thematisch, kwalitatieve informatie, en maatregelen voor de beveiliging van de dataset.
3. *Gegevensdefinitie* met een definitie voor de semantische, thematische geometrische en temporele eigenschappen van de objecten, conceptuele schema's voor de gegevensstructurering met objectclassificatie, verwijzingen naar externe thesauri.

Meestal worden metadata normen voorzien van een minimaal vereiste set van metadata elementen. Zo zegt de CEN ENv (Voornorm) 12657:1998, Geographic Information – Data description – Metadata dat tenminste een van de vermelde parameters van de metadata norm moet worden vermeld.

ISO19115: 2003, Geographic information – Metadata heeft een specifiek opgegeven minimale dataset die een opsomming is van de verplichte velden uit de complete metadataset.

Ontwikkelingen

De ontwikkeling van de ISO, Geographic information – Metadata norm was gebaseerd op studies van verschillende eerdere ontwikkelingen op het gebied van metadata normen, zoals:

- ANZLIC Metadata: Core Metadata Elements, 1995, toegepast in Australië en Nieuw Zeeland;
- Canadian Directory Information Describing Digital Geo-referenced Data Sets, 1994;
- CEN ENv 12657:1998, Standard for Geographic Information – Metadata, de Europese voornorm voor metadata;
- FGDC Content Standard for Geo-spatial Metadata, 1994, de metadata norm voor de VS;
- DIGEST, 1994 voor toepassing binnen de NAVO;
- IHO S 57, 1995, ontwikkeld voor hydrografische toepassingen;
- Dublin Core Metadataset. Deze metadataset is oorspronkelijk ontwikkeld voor de bibliotheken, maar nu ook veel toepassing vindt in uitgeverij en overheden.

Metadata bevraging en distributie

Er zijn vele mogelijkheden om metadata te verspreiden, waarbij Interent een belangrijk hulpmiddel is geworden. Bij het zoeken op Internet naar gegevensbronnen kan men drie niveaus onderscheiden:

1. *Metadata ontdekken.* Daarvoor is nodig een set van gemeenschappelijke termen, die de inhoud van de informatiebronnen beschrijven, een standaard grammatica om deze termen te kunnen samenvoegen in begrijpbare zinnen en een raamwerk dat gebruikers in staat stelt de metadata over te dragen. Samen vormen ze een architectuur voor de beschrijving en overdracht van metadata bronnen op het Internet. In de eenvoudigste vorm bestaat het systeem uit een centrale database met beschrijvingen van verschillende dataset, waarin iedere leverancier via speciale programmatuur zijn eigen gegeven kan bijhouden. Daarnaast bestaan er ook vormen waarin de metadata bij de leverancier blijft en via Internet toegang tot de beschrijvingen wordt verschaft.
2. *Metadata interpreteren.* Hierbij zijn de formele afspraken over de inhoud van de metadata van belang: gebruikers moeten een indruk kunnen krijgen van de inhoud van de dataset, waar de dataset is verkregen en de voorwaarden waaronder de dataset beschikbaar is voorbeelden hiervan zijn:
 - MEGRIN Geographical Data Description Directory service in Europe;
 - European Spatial Metadata Infrastructure (ESMI);
 - Asia-Pacific Spatial Data Infrastructure (APSDI) zoals voorgesteld door PC GIAP (Permanent Committee for Geographic Information in Asia en de Pacific);
 - Inter-American Geo-spatial Data Network (IGDN) in Latin America and the Caribbean;
 - National Geospatial Data Clearinghouse in the U.S.A., following the Executive Order 12906 van President W. Clinton van 4 april 1994;
 - Canadian Geospatial Data Infrastructure GEONet, ontwikkeld door GeoConnections;
 - Australian Spatial Data Infrastructure Distribution Network;
 - Japanse GSI clearinghouse gateway system met multi-byte karakter code sets uit het GEO profiel, die is gebaseerd op ISO 23950.

De meeste van de huidige normen bevatten metadata naast de geografische gegevens. Bijvoorbeeld CEN ENv12657,1998 Geographic Information – Data description – Metadata en de ISO 19115, 2001 Geographic Information – Metadata geven een beschrijving van gebruikte terminologie, metadata elementen en een schema, dat, indien juist toegepast, gebruikers in staat stelt gegevens te lokaliseren, toegang te verkrijgen, te evalueren, te selecteren, aan te schaffen en toe te passen op een efficiënte wijze. Bijkomend voordeel voor de leveranciers is dat ze de organisatie en faciliteiten van de gegevens voor hen ook regelen en ook de karakteristieke eigenschappen van de gegevens opsommen om ze te catalogiseren.

CEN ENv 12657 gebruikt EXPRESS en EXPRESS-G als modelleertaal, terwijl ISO19115 UML (Universal Modeling Language) toepast; ze zijn een integraal deel van de abstracte modellen voor geografische informatie, zijn computer leesbaar, hebben een grafische voorstelling en verschaffen een beeld van de klassentypen, attributen en het type en kardinaliteit van de relaties.

3. *Data verkrijgen en toepassen.*

Gebruik van metadata

In het gebruik van metadata zijn enkele opvallende voorbeelden bekend geworden (Ahonen-Raino, 2001), (McClland, 2002), die ons het volgende leren:

1. Meestal achten gebruikers de verstrekte gegevens juist, wat ten onrechte blijkt te zijn, omdat metadata incompleet zijn en fouten bevatten. Onverwachte toepassingen zoeken naar informatie die niet bestaat en dan denken gebruikers dat de metadata onvolledig is.
2. Gespecialiseerde woordenboeken bemoeilijken de gegevens invoer en het begrip. Het gebruik van metadata door leken leidt vaak tot een Lego™ methode van bouwstenen en kinderen zijn niet bevooroordeeld door de bedoelingen van de leveranciers: zij passen dezelfde ontwerpen toe in zowel een waterrijke omgeving als in de woestijn. Leken kunnen dezelfde filosofie toepassen binnen een geografische omgeving.
3. Het is gebruikelijk dat normen tegelijk worden ontwikkeld met de toepassingen; dat geldt ook voor metadata. Conceptuele modellen zouden deze nieuwe toepassingen moeten volgen: standaarden zijn daarom ook nooit af!

Vanuit de gebruiker gezien is het belangrijk te weten hoe goed de garantie van de producent is voor de metadata informatie; dit is ook van belang voor de producenten omdat onjuiste metadata informatie hen duur kan komen te staan. De praktijk laat een tendens zien dat klanten metadata evalueren in de volgende volgorde, bedekking, actualiteit en volledigheid en kwaliteit (Jakobsson, 2002).

De Werkgroep Kwaliteit in CERCO (Comité Européen des Responsables de la Cartographie Officielle) heeft in 1998 een questionnaire gemaakt, die werd toegezonden aan de 32 nationale karteringsorganisaties in Europa. De resultaten zijn geheel beschikbaar op de website van EuroGeographic – <http://www.eurogeographics.org> en er kunnen enkele interessante conclusies uit worden getrokken:

- Gebruikerseisen worden door interviews met de klanten vastgesteld.
- De meeste datasets hebben een product specificatie, die is gebaseerd op 'in-house' ontwikkelde methodologie omdat internationale normen nog niet beschikbaar waren (nu ISO9115), hoewel enkele datasets op internationale productspecificaties als Corine of Digest, enz. waren gebaseerd. Productspecificaties omvatten:
 - gegevenswoordenboeken;
 - inhoudspecificaties;
 - inwinningspecificaties;
 - technische specificaties;
 - gebruikershandleidingen.

Regionale en nationale metadata ontwikkelingen

In verschillende regio's in de wereld zijn specifieke ontwikkelingen ontstaan uit een behoefte naar metadata informatie.

Europa

De CEN ENv 12657:1998, Geographic Information – Data description – Metadata is alom geaccepteerd in Europa en in veel landen wordt de norm gebruikt voor metadata beschrijving. Pan-Europese projecten zoals Geographical Data Description Directory (GDDD), Geo-Scientific Electronic Information Exchange System (GEIXS), Added Value Information Dissemination voor hydrografische Datasets (AVID), European Spatial Metadata Infrastructure (ESMI), enz. passen CEN ENv 12657:2001 toe voor de distributie van gegevens.

Ook zijn er speciale projecten geweest die zich bezighielden met de veeltalige situatie in Europa, zoals Electronic Trade for Geographic Information (GISED), Methods for Access to Data and Metadata in Europe (Madame) en CLEF (Cross-Language Evaluation Forum). Het Open

Archive Forum ondersteunt projecten en nationale initiatieven met een open archief benadering als een interoperabel raamwerk met Open Archives Initiative (OAI) metadata protocol in Europa.

Door het opstarten van het Europese project INSPIRE in het zesde Framework Programme met de noodzaak voor de ontsluiting van vele Europese bestanden is een harmonisatie tussen de ISO19115:2003 en CEN ENv 12657:1998 nodig gebleken. Daarvoor is nog een formele weg af te leggen: CEN/TC 287 in het voorjaar van 2003 weer in het leven geroepen en de eerste stappen voor dat proces zijn al ingang gezet.

Noord Amerika

Tussen Canada en de V.S. hebben informele discussies geleid tot een MoU (Memorandum of Understanding), die ertoe moet leiden dat ANSI/INCITS L1 en SCC gezamenlijk een profiel van ISO191xx normen gaat ontwikkelen. Momenteel wordt in de V.S. de FGDC Metadata Content Standard for Digital Geospatial Metadata, STD-001-1998 Version 2.0. veelvuldig gebruikt.

Azië/Pacific

Versie 1 van de ANZLIC Metadata Guidelines werd gepubliceerd in 1996 gevolgd door 'Recommended Guidelines for the Transfer of ANZLIC Metadata Core Elements'. Deze twee documenten zijn door de industrie goed ontvangen en worden veelvuldig toegepast. In het gebruikt zijn echter enkele inconsistenties en tekortkomingen gevonden, waardoor het nodig was enkele aanvullende metadata elementen te definiëren in 1999. Dit werd gezien als een interim maatregel omdat de internationale norm 'ISO 19115, Geographic Information – Metadata' op zich liet wachten.

De Aziatische regio is zeer divers en strekt zich uit van Armenië, Iran, Azerbeidjaan en Kazakstan in het Noordwesten, tot Nieuw Zeeland, Frans Polynesië, Samoa en Cook eilanden in het Zuidoosten. PC GIAP is de centrale organisatie (met 55 landen leden) en heeft een Data Node project in 2000 gestart met metadata als een centraal thema en ISO19115:2003 als basis.

Centraal end Zuid Amerika en het Caribische gebied

De regio is zeer actief in de ontwikkeling van metadata als onderdeel van verschillende SDI ontwikkelingen waarbij vier regionale organisaties een rol spelen: PCIDEA (Permanent Committee on Spatial Data Infrastructure for the Americas), PROCIG (Central American Development Project for GIS), PAIGH (Pan American Institute of Geography and History) and DIGSA (Directory of Geographic Institutes from South America, Spain and Portugal). De meeste landen in de regio passen de FGDC standaard toe, maar ook hebben ze veelal besloten om die door de ISO19115 te vervangen.

Afrika en het Midden Oosten

In deze regio gebeurt niet veel op de ontwikkeling van metadata, ofschoon in Zuid Afrika wel enkele toepassingen van ISO 19115 zijn te vinden. De meeste ontwikkelingen liggen in projecten van de United Nations Environmental Programme (UNEP) en de Regional Remote Sensing unit of the Southern African Development Community in Harare (SADC).

Conclusies

De toepassing van GIS en dus ook metadata zijn globaler en grensoverschrijdend geworden, wat is te zien in veel toepassingen op het gebied van milieu, bosbouw, mariene en culturele bescherming, rampen bestrijding, enz. De verschillende regionale ontwikkelingen hebben ertoe geleid dat een internationale norm voor metadata ISO 19115: 2001, Geographic Information – Metadata is ontwikkeld. De meeste landen nemen zich voor om norm in te voeren, hoewel in Europa en dus ook in Nederland eerst nog het formele traject voor Europese normalisatie moet worden doorlopen.

Literatuur

(Ahonon-Raino, 2001)

Description of the content of geographic datasets,

Ahonon-Raino, P., (National Land Survey, Finland) in Bjørke, J.T., Tveite, H. (eds.) Proceedings of the 8th Scandinavian Research Conference on Geographical Information Science, 25-27, June 2001, Ås, Norway

(DLC, 1979)

Technical (Unclassified) Report UG478.A88, 1979

Defense Technical Information Center, Defense Logistics Center, U.S.A.

(Duval, 2002)

Metadata principles and practicalities

Duval, E., (KU Leuven, Belgium), W. Hodgins, (Autodesk), S. Sutton, (University of Washington), S.L. Weibel, (DCMI), D-Lib Magazine April 2002, Vol.8, no. 4, ISBN 1082-9873.

(Homer, 1978)

A mathematical model of the flow of data in a management information system

Homer, E.D., 11th American Meeting of the Institute of Management Sciences, 1978.

(ISO, 1946)

International Organisation for Standardisation. Contact: www.iso.ch

ISO/TC 211 Geographic information/Geomatics. Contact: <http://www.isotc211.org/>

(Jakobsson, 2002)

Data Quality and Quality Management – Examples of Quality Evaluation Procedures and Quality Management in European National Mapping Agencies

Antti Jakobsson in *Spatial Data Quality* Wenzhong Shi (HK Polytechnic University), Peter F. Fischer (University of Leicester) and Michael F. Goodchild (University of California) Eds., Taylor and Francis London UK, 2002 ISBN 0-415- 25835-9

(McCelland, 2002)

Challenges for service providers when importing metadata in digital libraries

M. McCelland, D. McArthur, S. Giersch (Colleges Eduprise), G. Geisler (University of North Carolina), D-L lib Magazine April 2002, Vol.8, no. 4, ISBN 1082-9873.

(McIntosh, 1968)

Information Processing 68

S. McIntosh, D. Griffel, Proceedings of International Federation for Information Processing (IFIP) Congress, August 1968.

(Moellering, 1991)

Spatial database transfer standards: current international status

Harold Moellering, editor. Published on behalf of the International Cartographic Association by Elsevier Applied Sciences, London 1991. ISBN 185166677X.

(Moellering, 1994)

Technical characteristics for assessing standards and for the transfer of spatial data and brief international descriptions.

Harold Moellering and C. Clement, (Eds.). International Cartographic Association Standards Commission for the transfer of spatial data, Columbus, Ohio, U.S.A., 1994, 104 pp.

(Moellering, 1997)

Spatial database transfer standards 2: characteristics for assessing standards and full description of the national and international standards in the world

Harold Moellering, editor, Richard L. Hogan associate editor. Published on behalf of the International Cartographic Association by Elsevier Applied Sciences, Oxford, UK and Tarrytown, N.Y. 1997. ISBN 0 08 042433 3.

(Schelling, 1978)

The use of IBM's data dictionary

G. Schelling, Proceedings of the Conference on Data dictionary Systems, London, Nov 1978

(Weber, 1978)

Data Base theory and practice

H. Weber, A.I. Wasserman, Conference in Berlin, 13-15 Sept. 1978.

(Williamson, 2003)

Spatial Data Infrastructures – From Concept to Reality

Ian P. Williamson, Abbas Rajabifard, Mary-Ellen F. Feeney (eds) Taylor and Francis 2003.

(Uitermark, 2001)

Ontology-Based Geographic Data Set Integration

H. Uitermark, PrintPartners Ipskamp Enschede, The Netherlands ISBN 90-365-1617-X, 2001.

(Ziegler, 1978)

Distribution: A New Impetus Toward Understanding Data

K. Ziegler, (International Business Machines, Inc.), August 1978.

Organisatie	Voor on-line contact zie URL:
Dublin Core Metadata Initiative	http://www.xml.com/pub/2000/10/25/dublincore/
Institute for Electrical and Electronic Engineers, Learning Object Metadata Working Group	http://grouper.ieee.org/p1484/doc/wg12/LOM-WD3.htm
U.S.A National Spatial Data Infrastructure	http://www.fgdc.gov
FGDC Official statement, Department of the Interior, Office of the Solicitor	http://www.fgdc.gov/metadata/meta_trademark.htm
International Organisation for Standardisation	http://www.iso.ch
ISO/TC 211 Geographic information/Geomatics	http://www.isotc211.org/
EuroGeographic	http://www.eurogeographics.org

Samenvatting

Vroeger werden kaartgegevens op papier bij de gebruikers afgeleverd, die op zijn beurt in transparante overlays eigen informatie kon toevoegen of met de gegevens op de kaart analyses uitvoeren. Nu we in een digitaal tijdperk zijn aangeland worden beschrijvingen van kaartbestanden in metadataseten geleverd, die via Internet beschikbaar kunnen worden gesteld. Door de globalisering is een uniforme syntactische en semantische formulering noodzakelijk zodat gebruikers uit verschillende culturen wel dezelfde interpretatie geven aan de metadata.

Daarbij zijn verschillende ontwikkelingen van belang geweest sinds de ingebruikname van de metadata beschrijvingen, zoals regionale, nationale en toepassingsgerichte definities van meta-

data normen. Verschillende theoretische en organisatorische ontwikkelingen zijn daarbij in acht genomen.

Tot slot wordt de huidige stand van zaken m.b.t. het gebruik van metadata in verschillende regio's in de wereld toegelicht.

GeoKey - de sleutel tot uw 'Map Services'

Ing. Ronald C. Bottelier
Geodan IT BV Amsterdam
ronald.bottelier@geodan.nl

Inleiding

Over meta-informatie van geografische bestanden en de toegevoegde waarde voor organisaties is al veel geschreven. Over het wat en waarom zal dit artikel *niet* gaan. *Wel* over de mogelijkheden die nieuwe standaarden en de technieken anno 2004 bieden om uw ruimtelijke gegevens organisatiebreed te ontsluiten en hoe het product GeoKey van deze standaarden gebruik maakt. Een blik op de toekomst, het heden en het verleden van GeoKey in omgekeerde volgorde.

GeoKey Flashed Back

De start

GeoKey is sinds 1995 het meta-informatie systeem van Geodan. Ontstaan vanuit de behoefte bij diverse klanten tot het vastleggen en opvragen van meta-informatie. Immers, in deze periode groeide het aanbod van geografische gegevens stormachtig, en daarmee groeide de behoefte tot een efficiënte, centrale beschrijving en beschikbaarstelling van de metadata.

In 1995 was de situatie anders dan nu: er bestond nog geen standaard voor de opslag van meta-informatie, XML was nog niet uitgevonden en internettoepassingen stonden in de kinderschoenen. Hierdoor waren de eerste versies van GeoKey gebaseerd op een door Geodan ontwikkeld datamodel en meta-informatie formaat. De wijze van opslag en raadpleging van de metadata was in nauwe samenwerking met de klantenkring opgesteld en voldeed daardoor prima. GeoKey was toen het enige beschikbare product voor geografische meta-informatie op de markt en dichtte daar ook een *gat*.

Met een toename van het gebruik van informatie komt vaak de behoefte tot onderlinge uitwisseling en daarmee de roep tot standaardisatie.

Standaard CEN98

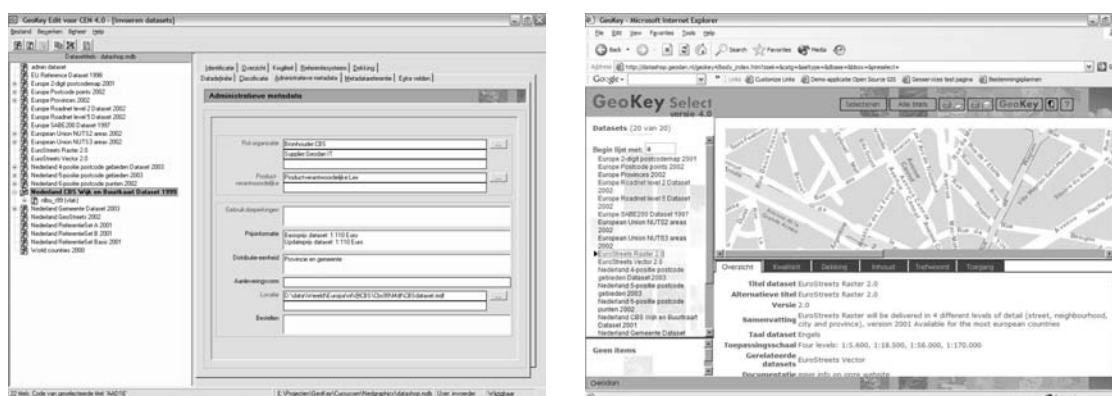
De eerste ontwikkeling in de standaardisatie van meta-informatie was de introductie van de CEN98 norm, de Europese norm voor de opslag van meta-informatie. Deze norm werd als standaard geadopteerd bij Rijkswaterstaat en LNV. Het antwoord van Geodan was het introduceren van een nieuwe GeoKey versie gebaseerd op CEN98. Het datamodel werd aangepast en de CEN98 hoofdstukindeling was in de gebruikersinterface van GeoKey terug te vinden. In onze inspanning om GeoKey op CEN98 aan te sluiten was de begrijpelijke reactie van veel gebruikers: "Wat veel velden!", "Moet ik die allemaal invullen" en "Wat betekenen die allemaal?".

Een metadata standaard kiezen is één, de standaard vervolgens invoeren en toepassen is twee en vereist begrip, discipline en kennis van de meta-informatie invoerders: Vandaar dat de implementatie van GeoKey en CEN98 in een organisatie altijd wordt gecombineerd met een interactieve sessie met de beoogde gebruikers. In deze sessie wordt de CEN98 standaard vertaald naar begrippen en werkwijze van de betreffende organisatie. GeoKey maakt CEN98 op maat om dit

zo goed mogelijk bij de gebruikers te laten aansluiten. Deze benadering is succesvol en maakt dat de standaard hanteerbaar blijft. CEN98 is nog steeds de basis van het huidige GeoKey.

GeoKey 4.0

In het kort een beschrijving, de functionaliteit en de toegepaste techniek van GeoKey 4.0. GeoKey 4.0 bevat veel op gebruikerswensen gebaseerde nieuwe functies, zowel in de invoer- en beheermodule GeoKey Edit, als in de raadpleegmodule GeoKey Select.



Figuur 1. Gebruikersinterface GeoKey Edit en GeoKey Select.

GeoKey Edit is gebaseerd op het beschrijven van fysieke gisbestanden, zoals SHAPE of TAB bestanden en van Oracle Spatial tabellen. Van deze bestanden kan de zogenaamde inherente metadata worden uitgelezen. Dit zijn bijvoorbeeld de geografische locatie van de data, het aantal objecten en de aanwezige attribootvelden. Overige metadata moet zelf door de beheerder worden ingevuld. Metadata kan worden geïmporteerd en geëxporteerd van en naar XML. Ook kan een XML worden aangemaakt voor gebruik in een node van het Nationaal Clearinghouse Geo-Informatie (NCGI), waardoor GeoKey een prima beheermodule is voor organisaties die hun metadata ook via het NCGI beschikbaar stellen.

De meta-informatie wordt opgeslagen in een Microsoft Access database, waarop de beheerder in GeoKey Edit gebruikersrechten kan toekennen. Dit maakt het mogelijk om de beschikbaarstelling *gericht* te doen. De beheerder kan zo nauwkeurig bepalen welke medewerkers of afdelingen welke meta-informatie mogen raadplegen. Opslag in een database maakt het beheer eenvoudig en maakt het mogelijk om deze centraal te ontsluiten en bevragen.

GeoKey Select verzorgt de ontsluiting van de GeoKey metadatabase(s) via het intranet of internet. Zoals dat met ruimtelijke gegevens vanzelfsprekend is, kan gezocht worden op administratieve kenmerken (*zoekterm* 'geodesie') of *ligging* (binnen een aan te geven zoekgebied in een kaart). Het zoekresultaat wordt *live* in een internet webmapping scherm getoond of als plaatje, uiteraard met bijbehorende tekstuele meta-informatie. In het geval de gebruiker van de beheerder de juiste rechten heeft gekregen, kunnen gevonden gisbestanden worden gedownload voor gebruik in het eigen -stand alone- GIS.

GeoKey Select kan gedistribueerd werken, dat betekent dat via één GeoKey Select webapplicatie meerdere GeoKey database servers zijn te benaderen, bijvoorbeeld van meerdere afdelingen of vestigingen die zo elk hun eigen geografische data kunnen onderhouden en beschikbaar stellen.

Ook integratie van GeoKey met standaard GIS producten is mogelijk. Voor bijvoorbeeld ArcGIS en MapInfo Professional wordt de GeoKey extensie meegeleverd. Hiermee kunnen via de GeoKey Select server GIS bestanden opgezocht worden in de GeoKey database en kunnen de gevonden gisbestanden *direct* worden ingelezen in het GIS systeem voor verder gebruik.

GeoKey 4.0 is een krachtig product om geografische data organisatiebreed te ontsluiten. Maar in het land van metadata standaarden en technieken voor ontsluiting waait een nieuwe wind

GeoKey Flashed Foreward - Metadata in de services architectuur

De evolutie van de geo-ICT heeft zich ontwikkeld van monoliete gesloten systemen, via client-server technologie naar de actualiteit van de web services architectuur. De web services architectuur is, zoals de naam al zegt, volledig gebaseerd op internet technologie. Het internet staat inmiddels centraal in veel organisaties als hét medium voor gegevensuitwisseling.

In een architectuur gebaseerd op services kunnen web applicaties worden samengesteld die verschillende *diensten* verzorgen, bijvoorbeeld adreslokalisering, routeplanning en het serveren van achtergrondkaarten. Door diensten aan elkaar te koppelen, kunnen flexibel applicaties worden samengesteld. De kleinste eenheden programmatuur in deze architectuur worden *componenten* genoemd.

Het volgende is nu denkbaar: een internet web mapping applicatie bevat een component voor het zoomen naar adressen, een ander component om geografische selecties te maken en een component die de kaartlagen inlaadt en beheert. Een prachtige architectuur waar prima plaats is voor metadata services om de gewenste kaartlagen te kunnen vinden. De metadata service als centrale component! Maar volgens welke standaard?

Metadata en Open standaarden

In de afgelopen 10 jaar is ook meta-informatie meegegaan in de vaart der standaardisaties. Zo is er nu naast CEN98 ook ISO19115 beschikbaar als wereldstandaard. Zo niet belangrijker, is het feit dat meta-informatie is opgenomen in het Basic Services Model van het OpenGIS consortium (OGC) onder het begrip Catalog Services (WCAS of het recentere CS-W). Dit is een belangrijke ontwikkeling, omdat hiermee de ontsluiting van geografische data ook technisch wordt gestandaardiseerd in een services architectuur. De OpenGIS Web Catalog Services (WCAS) komen hiermee op een gelijkwaardige lijn te staan met de al bestaande OpenGIS specificaties voor Web Mapping Services (WMS) en Web Feature Services (WFS).

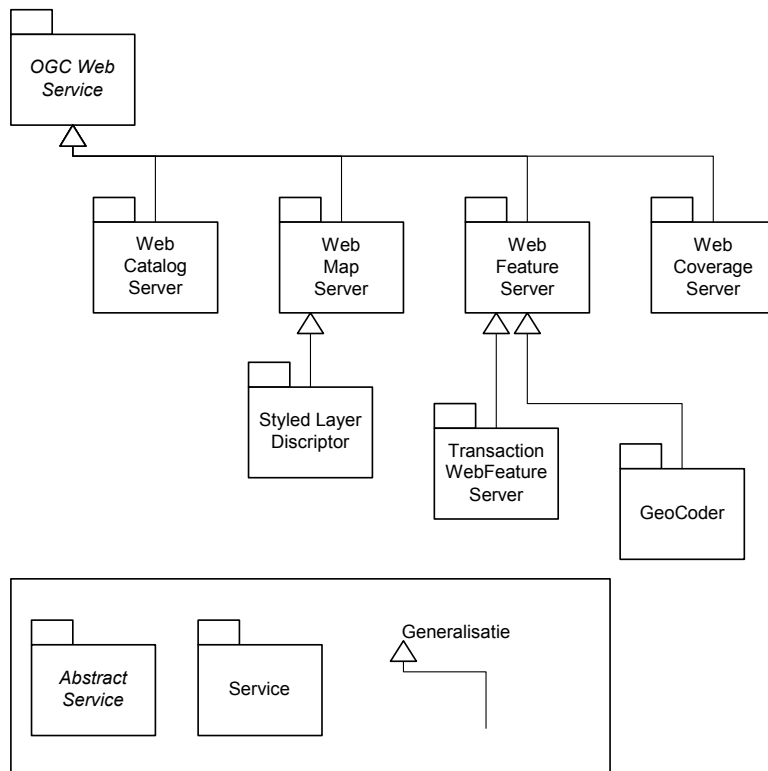
Metadata van map services

De introductie van web mapping heeft tot gevolg dat voor een gebruiker in een internet omgeving de fysieke locatie van (GIS-)bestanden minder relevant wordt. De locatieaanduiding verschuift immers van een padverwijzing naar een gedeelde map op het netwerk naar een URL, bijvoorbeeld 'g:\geopakhuis' wordt <http://...> De URL vertelt op welke server in het internet de gegevens zijn op te halen. Voor gebruikers wordt hiermee de beschikbaarheid van de kaart(laag) van groter belang dan de werkelijke fysieke locatie van het achterliggende gisbestand.

Ook wordt geo-informatie steeds vaker in ruimtelijke databases geïmplementeerd, zoals in Oracle Spatial. In een database is eigenlijk geen sprake meer van één fysiek gisbestand, maar van een geografische tabel, al dan niet met gerelateerde tabellen. Tenslotte is door de web mapping technologie ook het fysiek uitwisselen van ruimtelijke informatie in een ander licht komen te

staan. Het fysiek uitwisselen door gisbestanden aan elkaar op te sturen wordt vervangen door gewenste bestanden eenvoudig in de eigen webmapping applicatie transparant via OpenGIS te projecteren als nieuwe kaartlaag. De fysieke locatie van het gisbestand kan zich hierbij in principe overal ter wereld bevinden.

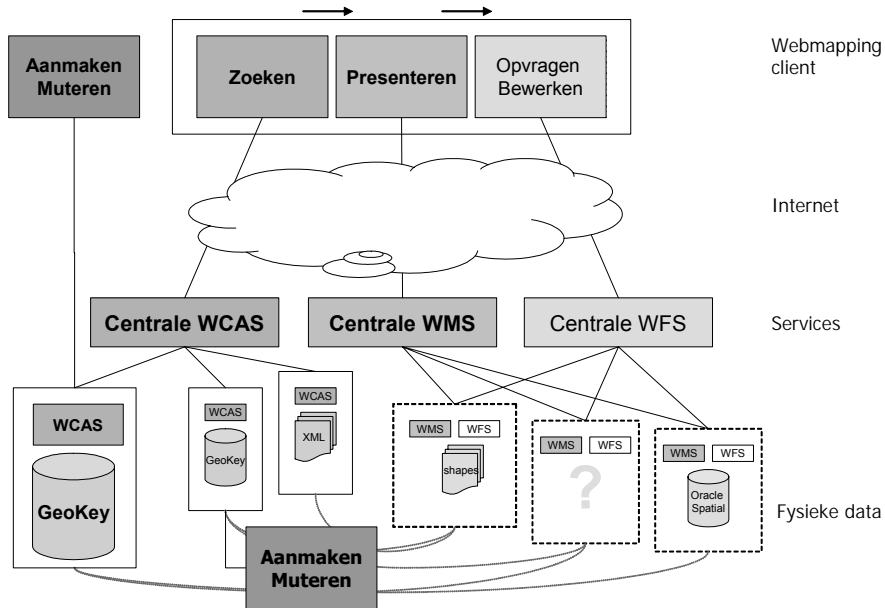
Dit alles vereist voor GeoKey een andere benadering van de wijze van ontsluiting van geografische bestanden: De verschuiving ontstaat van het denken in fysieke gisbestanden naar het denken in kaartlagen c.q. map services, alleen welke map service serveert deze kaartlagen?



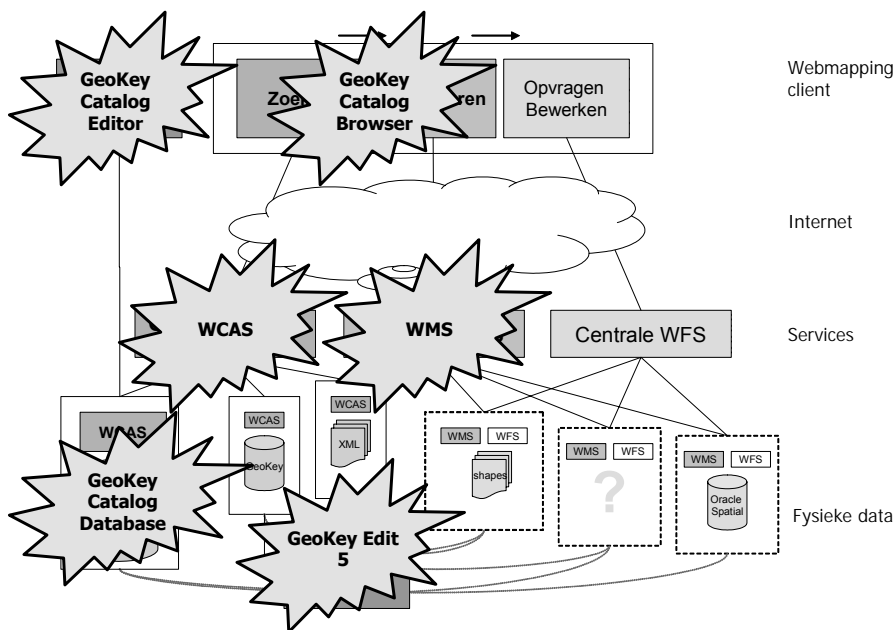
Figuur 2. Generalisatie OGC Basic Services Model.

GeoKey 5.0

GeoKey 5.0 is de werktitel voor de nieuwste GeoKey versie die volledig gebaseerd is op componenten en op open standaarden. Met GeoKey 5.0 haalt een organisatie meer een 'koffertje' met componenten in huis dan een kant en klaar product. De componenten kunnen naar wens worden ingezet. In het koffertje kunnen mogelijk ook open source componenten een plaats krijgen. GeoKey 5.0 wordt een OGC Web Catalog Server voor het zoeken van de metadata, het krijgt ISO19115 als standaard voor de opslag en uitwisseling en een OGC Web Mapping Server voor de presentatie van de gevonden kaarten in de browser. Uiteraard gaat de uitwisseling van metagegevens via XML en de uitwisseling van het kaartbeeld via GML, zoals het OGC dat voorschrijft. Overigens wordt ook een start configuratie meegeleverd die bestaat uit een aantal modules. In onderstaande illustratie staat de services architectuur, met daarin gepositioneerd de nieuwe GeoKey modules.



Figuur 3. GeoKey 5.0 modules in de services architectuur (1).



Figuur 4. GeoKey 5.0 modules in de services architectuur (2).

GeoKey Catalog Editor

Dit wordt de nieuwe invoermodule. Via de browser kunnen ISO19115 velden worden ingevuld en zo kaartlagen worden beschreven. De invoerder kan zelf kiezen of er één of meerdere kaartlagen worden beschreven als één meta-informatie entiteit. In de GeoKey database op de server wordt de kaartlaag (*layer* in OGC termen) opgeslagen als een URL. De database kan overigens Access of Oracle zijn.

GeoKey Catalog Browser

Dit wordt de basis webclient waarmee meta-informatie gezocht kan worden. Zoekwoorden kunnen worden ingevuld en het gevonden resultaat wordt met een WMS component als nieuwe kaartlaag getoond.

GeoKey WCAS Server

De GeoKey WCAS server wordt de schakel tussen de opslag van de fysieke metadata in databases of XML's en de gebruikers interface in de Catalog Browser. Ook kunnen hiermee GeoKey 4.0 databases benaderd worden.

GeoKey WMS Server

Dit component verzorgt de presentatie van de te tonen layers via OpenGIS aanroepen naar andere mapservers of de fysieke geografische data.

GeoKey Edit 5.0

Ook het al bestaande GeoKey Edit krijgt een plaats in het 'koffertje', echter de invoer blijft gebaseerd op CEN98.

Samenvattend

Met de komst van de OpenGIS Catalog Services specificaties heeft de opslag van metadata een volwaardige rol gekregen in de OpenGIS web services architectuur. Hierdoor is de *missing link* gelegd tussen mapservices, layers en hoe deze zijn te vinden. GeoKey 4.0 biedt al een prachtige basis voor meta-informatie, het toekomstige GeoKey 5.0 zal hiervoor een complete architectuur leveren die de basis kan zijn voor een volledig op mapservices gebaseerde ontsluiting van de geo-informatie in uw organisatie. De GeoKey componenten kunnen desgewenst in uw eigen bestaande services architectuur worden geïntegreerd en zo toegevoegde waarde leveren. De basisimplementatie van GeoKey 5.0 voldoet om de eerste stappen op het pad van OpenGIS web mapping in te slaan.

Uit ervaring weten we dat de meeste organisaties alleen de meest essentiële meta-informatie velden willen invoeren. De invoer in GeoKey 5.0 zal daarom compact blijven, maar wel gebaseerd zijn op ISO19115, en uitbreidbaar, c.q. aanpasbaar zijn aan de wensen van de individuele organisaties. Dit houdt de inspanningen voor invoer binnen de perken en maakt meta-informatie toegankelijk.

File based of layer based: GeoKey blijft de sleutel tot uw geo-informatie!

Sleutelpositie voor meta-informatie en meta-informatiebeleid in de Geo-Informatie Infrastructuur van Rijkswaterstaat

Dr. Michel J.M. Grothe en ir. Paul van Asperen
Adviesdienst Geo-informatie en ICT, Rijkswaterstaat
m.j.m.grothe@agi.rws.minvenw.nl p.vasperen@agi.rws.minvenw.nl

Inleiding

Binnen Rijkswaterstaat wordt sinds begin jaren negentig gewerkt aan een geo-informatie infrastructuur. Meta-informatie is daarbij vanaf het begin gezien als belangrijke peiler waarop de geo-informatie infrastructuur rust. Dit heeft in de jaren negentig in het zogenaamde 'GIS-huis' concept geresulteerd [1]. Meta-informatie werd daarbij - naast standaarden en afspraken - gezien als het 'cement' van het GIS-huis; de verbinding tussen de drie onderdelen van het huis: de geo-informatie (het fundament), de administratieve informatie (de bouwstenen) en de producten van de geo-informatievoorziening (het dak). Rijkswaterstaat heeft zich aanvankelijk wat betreft de meta-informatie voor de geo-informatievoorziening vooral gericht op het beschrijven van geodatasets. Zo zijn inmiddels ruim 40.000 geodatasets middels meta-informatie beschreven. Met de komst van geïntegreerde open webtechnologie voor geo-informatie, waardoor de geo-informatie sneller en makkelijker toegankelijk wordt, zal de rol van meta-informatie nog belangrijker worden. Dit wordt onderstreept in de in 2002 door TNO uitgevoerde Geo-ICT trendrapportage van Rijkswaterstaat [2], waarin een veranderende vraag naar geo-informatie is geconstateerd. Geo-ICT is op steeds meer plekken aanwezig, er komen nog steeds nieuwe toepassingen en er is een duidelijke vraag naar meer kwaliteit. Daarbij is in alle onderdelen van de informatieketen samenwerking tussen publieke en private partijen aan de orde. De onderlinge afhankelijkheden zijn daarbij enorm toegenomen om de gewenste kwaliteit aan informatie te kunnen leveren. Afnemers stellen aanzienlijk hogere eisen wat betreft snelheid en eenvoud van de geo-informatievoorziening. Ook mede omdat technologie dit mogelijk maakt [3]. Meta-informatie heeft daarbij een sleutelpositie in het snel en op maat aanleveren van geo-informatie. Meta-informatie is daarmee cruciaal om aan de geconstateerde veranderende vraag tegemoet te komen.

Dit artikel zal nader ingaan op de 'Meta-Informatie Infrastructuur' van Rijkswaterstaat. Richtinggevend voor de inrichting van deze infrastructuur is het in 2002 geformuleerde meta-informatiebeleid. Allereerst wordt de meta-informatie infrastructuur binnen Rijkswaterstaat omschreven en zal het meta-informatiebeleid worden geïntroduceerd. Daarna wordt in meer detail ingegaan op de onderdelen en invulling van de meta-informatie infrastructuur voor de geo-informatievoorziening van Rijkswaterstaat, te weten beleid en afspraken, standaarden, gegevens, systemen en technologie en de organisatie. Tot slot volgen enkele resumerende opmerkingen.

Meta-Informatie Infrastructuur

Sinds 1996 is Rijkswaterstaat bezig haar Meta-Informatie Infrastructuur (MII) in te richten en uit te bouwen. Binnen de geo-informatievoorziening van Rijkswaterstaat wordt meta-informatie eenvoudig omschreven als 'informatie, die informatie beschrijft'. In relatie tot geo-informatie betreft het informatie, die op een hoger abstractieniveau de geo-informatie beschrijft, bijvoorbeeld bron, nauwkeurigheid, et cetera. Meta-informatie is daarbij synoniem met metadata en metagegevens. De meta-informatie infrastructuur voor de geo-informatievoorziening binnen

Rijkswaterstaat is gedefinieerd als "het geheel aan beleid, afspraken, standaarden, gegevens, technologie en kennis over meta-informatie, die de gebruiker(s) ten dienste staan om zijn/haar eigen taken met behulp van geo-informatie uit te voeren" [4]. Daarbij dient tegelijkertijd te worden opgemerkt, dat de meta-informatie infrastructuur een moeilijk te definiëren begrip c.q. concept is. Het dient beschouwd te worden als een gestuurd en een zich spontaan ontwikkeld proces, dat zich geleidelijk voltrekt in nauwe samenwerking met het Nederlandse en internationale geo-werkveld. Door invulling van de essentiële bouwstenen van de meta-informatie infrastructuur zal deze geleidelijk en op iteratieve wijze haar vorm krijgen. Het gaat daarbij om die zaken van de meta-informatievoorziening, die voor verschillende gebruikers van gemeenschappelijk belang zijn. Het gaat niet om organisatiespecifieke aspecten, maar om zaken waarbij afstemming en samenwerking gezocht wordt voor een betere en goedkopere eigen taakuitoefening. Daarbij is de Meta-informatie infrastructuur geen doel op zich, maar staat ten dienst van de doelstellingen van Rijkswaterstaat en het gehele Ministerie van Verkeer en Waterstaat (VenW).

Om de ontwikkeling en implementatie van de MII te kunnen stroomlijnen en in te kaderen is in 2003 een beleidsplan voor meta-informatie gemaakt [4]. In het meta-informatiebeleid staat een 'lange-termijn' visie centraal. Deze visie - geformuleerd voor de periode tot 2007 - luidt als volgt: "Geo-informatie zonder metadata bestaat voor de organisatie niet meer. Er is VenW-brede toegang tot geo-informatie via de meta-informatie, en andersom, voor iedereen waarvoor dat nodig is". Ter realisatie van deze visie is tevens een 'Missie MII 2002-2007' gedefinieerd, die de implementatie van de visie operationaliseert: "In 2007 heeft iedere VenW medewerker ongeacht zijn locatie - thuis, op kantoor en in het veld - de beschikking over een intelligent zoekstelsel om 80% van alle binnen VenW aanwezige geo-informatie te vinden en op te vragen". Om de visie MII te kunnen realiseren zijn diverse beleidsuitgangspunten onderscheiden. De beleidsuitgangspunten zijn bedoeld als leidraad bij de invulling van de activiteiten, die nodig zijn om de visie te kunnen realiseren. Enkele voorbeelden van die activiteiten zijn:

- het organiseren van toegang tot de metadata van VenW voor zowel interne als externe organisaties;
- het hanteren van standaardcontracten voor het uitleveren van geo-informatie binnen VenW en het hanteren van een standaard dataparagraaf in projectplannen waarin is bepaald dat geo-informatie van metadata moet zijn voorzien;
- het opstellen van richtlijnen voor het aanleveren van metadata door derden zoals advies- en ingenieursbureaus. Voor verkenningen en planstudies zijn dergelijke richtlijnen reeds opgesteld [5].
- het ontwikkelen van een applicatiearchitectuur voor de MII volgens het 'open systeem' concept door toepassing van een gelaagde architectuur om integratie met andere systemen mogelijk te maken.

In de volgende paragrafen worden de diverse aspecten van het meta-informatiebeleid betreffende de geo-informatievoorziening en de invulling daarvan binnen Rijkswaterstaat verder beschouwd.

Beleid en afspraken

Aan de basis van de activiteiten, die Rijkswaterstaat onderneemt voor de ontwikkeling van haar MII staan een tweetal beleidsuitspraken. Deze beleidsuitspraken zijn bedoeld als leidraad bij de invulling van de activiteiten, die nodig zijn om de Missie MII 2002-2007 te kunnen realiseren. De beleidsuitspraken zijn hieronder verder toegelicht aan de hand van de concrete uitgangspunten en uitspraken, die tezamen het meta-informatiebeleid vormen. De eerste beleidsuitspraak luidt: "Geo-gegevens zonder meta-informatie bestaan voor de organisatie niet meer". Het twee-

de beleidsuitgangspunt luidt: "Er is VenW-brede toegang tot gegevens via de meta-informatie, en andersom, voor iedereen waarvoor dat nodig is". Het belang van metadata is groot en het gebruik van metadata is niet meer vrijblijvend. Gegevensbestanden kunnen niet worden uitgewisseld zonder metadata. Metadata bepaalt nl. mede de kwaliteit van de geo-informatie. Een gegevensbestand zonder metadata wordt beschouwd als niet-bestaand. Metadata en meta-informatie zijn integraal onderdeel van het gegevensbeheer. Deze uitgangspunten laten zich naar de dagelijkse praktijk als volgt vertalen in maatregelen [4]:

- De meta-informatie infrastructuur staat niet op zich maar maakt deel uit van de geo-informatie infrastructuur.
- De uitwisseling en toegankelijkheid van meta-informatie en geo-informatie binnen VenW vindt op een effectieve en efficiënte wijze plaats.
- Geen geo-informatie zonder metadata bij gegevensuitwisseling en -overdracht, zowel binnen VenW als aan en van derden.
- De toegang tot de metadata van VenW is zowel voor intern als externe organisaties geregeld. Toegang tot geo-informatie is mogelijk vanaf iedere locatie en op ieder moment, via het gebruik van standaard Internettechnologie, met en via krachtige zoekmachines en intelligente user interfaces.
- Voor het uitleveren van geo-informatie worden binnen VenW standaardcontracten gehanteerd. Tevens zijn projectplannen zoveel mogelijk met een standaard dataparagraaf over de levering van geo-informatie en metadata voorzien. Daarvoor zijn duidelijke standaardrichtlijnen aanwezig binnen VenW.
- VenW-brede basisbestanden worden door het Geo-Loket van de AGI gedistribueerd en van meta-informatie voorzien.
- Er zijn duidelijke richtlijnen voor het aanleveren van metadata door derden zoals advies- en ingenieursbureaus.

Naast deze meer algemene beleidsuitspraken ten aanzien van de meta-informatievoorziening zijn ook meer specifieke uitgangspunten geformuleerd ten aanzien van de het gebruik van standaarden, metagegevens, metasystemen en -applicaties en de organisatie. Deze zullen in de loop van dit artikel aan de orde komen in relatie tot de invulling daarvan binnen Rijkswaterstaat.

Standaarden

Rijkswaterstaat heeft standaardisatie hoog in het vaandel staan en heeft voor het vaststellen van standaarden een IT-Raad en Standaardisatieboard van Rijkswaterstaat. Het meta-informatiebeleid m.b.t. gebruik van standaarden luidt: "De meta-informatie infrastructuur maakt gebruik en sluit aan bij de ICT- en geo-ICT standaarden van Rijkswaterstaat". Daarbij zijn de volgende afgeleide uitgangspunten geformuleerd m.b.t. het adopteren van standaarden [4]:

- de MII volgt het I&A beleid van Rijkswaterstaat en de vastgestelde standaarden;
- de MII is gebaseerd op het concept van open systemen: de MII-tools voldoen aan de bijbehorende lagenarchitectuur zodat optimale mogelijkheden voor interoperabiliteit, portabiliteit en systeemintegratie is gewaarborgd.

Voor standaarden betreffende de meta-informatie in de geo-informatievoorziening richt Rijkswaterstaat zich momenteel op de volgende standaarden:

- het beschrijven van geodatasets conform NVN-ENV 12657 en ISO 19115;
- het beschrijven van services conform ISO 19919;

- het ontsluiten van bovenstaande meta-informatie conform OpenGIS Web Catalog Services (WCAS).

De belangrijkste standaarden en de wijze waarop Rijkswaterstaat deze heeft geadopteerd en geïmplementeerd, worden hieronder kort toegelicht.

NVN-ENV 12657 en VenW CEN98

Om meta-informatie onderling uit te wisselen is het van belang dat de meta-informatie op een uniforme wijze wordt beschreven. Binnen Rijkswaterstaat is de meta-informatie infrastructuur voor geo-informatie gebaseerd op Nederlandse en internationale standaarden en normen voor het beschrijven van geodatasets. In Nederland geldt hiervoor de Nederlandse voornorm NVN-ENV 12657, identiek aan de Europese CEN-norm uit 1998. Vaak wordt dan ook gerefereerd aan de CEN98-norm. Deze norm kent 290 elementen, het is praktisch onmogelijk om alle 290 elementen daadwerkelijk te gaan gebruiken. In 1999 heeft Rijkswaterstaat voor het Ministerie van Verkeer en Waterstaat het eerste Richtlijnendocument VenW CEN98 opgesteld. Voor het opstellen van de VenW CEN98 standaard voor het beschrijven van geodatasets is een selectie van de velden uit CEN98 gemaakt en is bovendien per veld beschreven op welke wijze dit veld moet worden ingevuld. Dientengevolge zijn ruim 80 velden uit de CEN98-norm opgenomen en 16 velden als organisatiespecifiek toegevoegd voor het beschrijven van aardobservatiebeelden. Eind 2003 is het Richtlijnendocument herzien (versie 3.0) en is door de centrale IT-Raad als standaard voor Rijkswaterstaat vastgesteld [6]. Naast Rijkswaterstaat wordt VenW CEN98 ook door diverse andere Nederlandse overheidsorganisaties gebruikt.

Van VenW CEN98 naar VenW ISO 19115?

Door ontwikkelingen in Europese context, waaronder de Europese Kaderrichtlijn Water, maar ook de samenwerking tussen ISO en het OpenGIS Consortium (OGC) op het gebied van OpenGIS web services (zie ook [7]), is in 2003 door Rijkswaterstaat een onderzoek gedaan naar omzetting van VenW CEN98 naar VenW ISO 19115. Voor een groot gedeelte is de omzetting van de CEN-velden naar ISO 1 op 1 mogelijk, in overige gevallen moeten velden worden gesplitst dan wel samengevoegd. In het algemeen kan worden gezegd dat ISO strenger is dan CEN. CEN laat veel vrije tekstvelden toe, terwijl ISO vaker met getallen en vastgestelde codelijsten werkt.

Op de zeer korte termijn zal 'pragmatisch' met de ISO-overgang worden omgegaan: alleen de meest relevante velden zullen naar ISO worden vertaald en in een aparte ISO-database worden opgeslagen. Op de langere termijn zal Rijkswaterstaat de Nederlandse norm blijven volgen; de grote vraag is dan ook op welke wijze en wanneer de Nederlandse normcommissie zal besluiten de ISO-norm als NEN-norm vast te stellen.

Echter, vanwege de adoptie van de OpenGIS web services architectuur wordt binnen Rijkswaterstaat inmiddels gewerkt met OpenGIS Web Catalog Services voor meta-informatie, die op ISO 19115 zijn gebaseerd. Daarvoor is het al noodzakelijk gebleken om voortijdig naast de CEN-richtlijnen een (beperkte) ISO-vertaling te hebben. Rijkswaterstaat heeft daarvoor haar eigen (beperkte) profiel gebaseerd op de VenW CEN98 velden.

ISO 19119; meta-informatie voor OpenGIS services

Met de officiële oprichting van Open GIS Consortium in 1994 is OpenGIS een begrip geworden in de geo-wereld. OGC heeft daarbij de samenwerking met ISO gezocht en in het verlengde daarvan zijn enkele ISO-standaarden geadopteerd en (her)gebruikt (m.n. ISO TC 211 Geoma-

tics). OGC hanteert 2 niveaus van standaardisatie: 1. de abstracte of officiële standaarden en 2. implementatiestandaarden, die interface(protocollen) van (een deel) van de abstracte standaarden beschrijven. Voor meta-informatie over geodatasets heeft OGC de standaard van ISO TC211 geadopteerd (ISO 19115). Daarnaast is een standaard ontwikkeld voor het beschrijven van services (ISO 19119). OGC heeft namelijk al vrij snel gekozen om haar activiteiten te richten op een architectuur voor de ontsluiting van geo-informatie in en via open, gedistribueerde netwerken, zoals Internet. Dit heeft geleid tot het OpenGIS Service Framework, waarin interfaces en protocollen zijn opgenomen voor samenwerkende geografische diensten, die het publiceren, vinden, en opvragen van geo-informatie (zie [8]) mogelijk maken. Diverse OpenGIS Web Services specificaties zijn beschikbaar, zoals services voor het presenteren en transformeren van geo-informatie. Voor het beschrijven van geodatasets en web services is respectievelijk een specificatie voor de Web Catalog Service en de Web Registry Service (gezamenlijk aangeduid met WCAS) beschikbaar. Op de binnen Rijkswaterstaat geïmplementeerde Web Catalog en Registry Service wordt hieronder in het kader van de applicatiearchitectuur verder ingegaan.

Gegevens(beheer)

Rijkswaterstaat wordt in toenemende mate beoordeeld op de juistheid van beleid en uitvoering van haar taken. De meeste werkprocessen van VenW zijn sterk gegevensgericht. De gegevenslogistiek zal sterk op metadata (gaan) leunen en zal daarmee ook sterk afhankelijk zijn van de kwaliteit van de metadata. Metadata zorgt voor eenduidigheid in het beheer van gegevens en voor een stuk digitale duurzaamheid. Het vinden van het juiste bestand en de juiste versie is hier een voorbeeld van. Om optimale toegankelijkheid voor een brede groep eindgebruikers te kunnen garanderen op korte en lange termijn is efficiënt beheer van meta-informatie noodzakelijk. Daarnaast is beleid voor gegevensuitwisseling en beleid over de kwaliteit van metadata essentieel. Voor de uitvoering van dit beleid is het metagegevensbeheer verantwoordelijk. Op het gebied van het metagegevensbeheer is het beleid eenduidig geformuleerd: "Meta-informatie is van hoge kwaliteit". Het verbeteren van de metadata kwaliteit wordt o.a. met behulp van de volgende beleidsuitspraken tot stand gebracht [4]:

- De materiemensen vervullen een belangrijke rol bij het metagegevensbeheer door de metadata door hen te laten vullen. Zij zullen inhoudelijke inconsistenties en definities eerder opsporen. Indien dit problemen oplevert, neem de gegevensbeheerder in het projectteam op, die de rol van metagegevensbeheerder vervult.
- Bij een (externe) opdracht over het leveren van geodata worden duidelijke afspraken gemaakt tussen opdrachtnemer en opdrachtgever over het leveren van de meta-informatie conform de standaarden van VenW.
- Er is een optimale ontsluiting van metadata en geo-informatie om ook projectmedewerkers en projectleiders te laten zien wat VenW aan geo-informatie heeft, waardoor metagegevensverzamelingen beter gebruikt (en gevuld) worden.
- Rijkswaterstaat beschikt over een kwaliteitshandboek dat moet garanderen dat de metadata-base voor geo-informatie op een uniforme wijze en kwalitatief hoogwaardige wijze wordt beheerd.
- Het metagegevensbeheer maakt deel uit van het reguliere gegevensbeheer bij en van geo-informatie. Dit betekent meer aan de bron beheren en niet later toevoegen, wat inhoudt dat er een sterkere relatie tussen data en metadata moet zijn.
- Metadata wordt bij voorkeur decentraal verzameld door de inhoudelijk verantwoordelijke (bij de bron), met uitzondering van VenW brede basisbestanden.
- Elke organisatie is verantwoordelijk voor de eigen interne meta-informatie en is daarmee verantwoordelijk voor zijn knooppunt in de meta-informatie infrastructuur.

- Beter archiveren. Zorg dat metabeheerders de ruimte krijgen om projectmedewerkers op te sporen en aan te sporen om hun data te ontsluiten voor de toekomst via opname in het metadata-archief.
- Het metagegevensbeheer is niet complex: er zijn goede tools om metadata te genereren en te beheren.

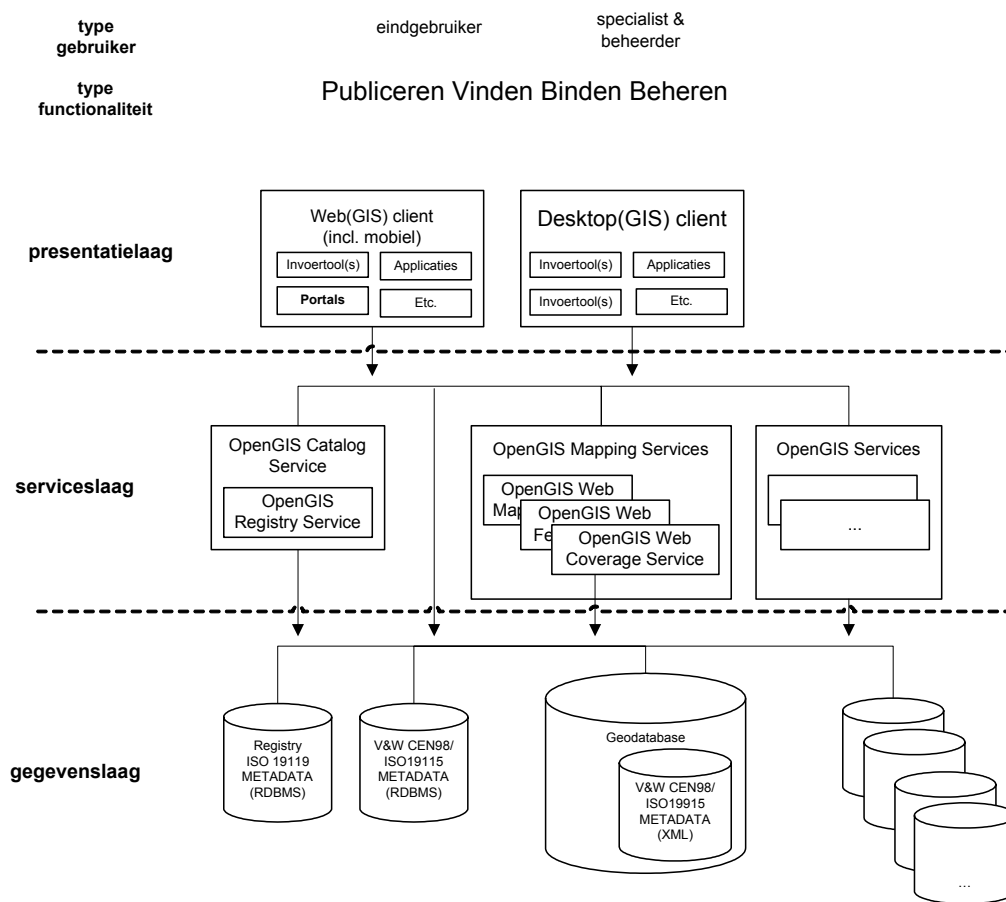
Technologie; de applicatiearchitectuur

Het beleid m.b.t. de technologie van de MII kan als volgt samengevat worden [4]:

- Geen geo-informatie zonder metadata in geo-systemen en geo-applicaties.
- Metadata en meta-informatie zijn een integraal onderdeel van de standaard opslagsystemen. Geen geo-informatie zonder metadata op de geoserver.
- De opslagsystemen genereren automatisch metadata: de inherente metadata.
- Streef naar eenvoud en samenhang in metasystemen.
- Metagegevensbeheer wordt zoveel mogelijk uitgevoerd m.b.v. marktpakketten.
- Verzorg de interfacing tussen systemen op een optimale wijze. Laat onderdelen met elkaar communiceren via open standaarden.
- Vastgelegd is hoe de keten van informatiesystemen functioneert.
- Vastgelegd is wie verantwoordelijk is voor het wijzigen van de keten van metasystemen en wie opdracht daarvoor zal geven.
- De MII is 'open' volgens het 'open systeem' concept en beschikt over een gelaagde architectuur om integratie met andere systemen mogelijk te maken.

De applicatiearchitectuur van de geo-informatievoorziening kent een viertal functionele uitgangspunten; de architectuur moet functies bieden voor het publiceren, vinden, binden (of opvragen) en beheren (inclusief be- en verwerken) van geo-informatie(bronnen). Dit kan niet zonder meta-informatie. Derhalve ziet de globale applicatiearchitectuur voor de inrichting van de geo-informatie infrastructuur binnen Rijkswaterstaat er als volgt uit (zie figuur 1). De architectuur is voor een belangrijk deel gebaseerd op de OpenGIS web services architectuur (zie [8]). De architectuur is gelaagd opgebouwd en bestaat uit een gegevenslaag, serviceslaag (ook wel applicatielaag) en presentatielaag. In de presentatielaag zitten zowel web clients als desktop clients met verschillende verschijningsvormen, zoals GIS-systemen, specifieke GIS-applicaties, meta-informatiesystemen en portals.

In de serviceslaag bevinden zich naast services voor kaartpresentatie en -bewerkingen (verschillende web servers voor verschillende typen geodata) de OpenGIS Web Catalog Services. Deze services zorgen ervoor, dat de gebruiker meta-informatie kan raadplegen en (gedistribueerd) opvragen. In de gegevenslaag zijn metagegevens opgeslagen conform de afgesproken standaarden (in diverse fysieke formaten). De eerder geschetste standaarden CEN98 en ISO 19115 voor geodatasets en ISO19119 voor services zijn daarbij momenteel leidend. De implementatie van deze standaarden kan diverse vormen aannemen; van filebased (in bijv. XML's) tot opslag in een RDBMS. Uitgangspunt is dat de Catalog Service toegang heeft tot de database of file-base voor het ophalen (zoeken, vinden en binden) van de geo-informatie. Gedistribueerde toegang is daarbij eveneens gewenst; er zijn namelijk meerdere (meta)databronnen aanwezig op verschillende fysieke locaties binnen (en buiten) Rijkswaterstaat. Daarnaast zullen voor het beheren van de metagegevens (toevoegen, verwijderen en wijzigen van metagegevens) ook rechtstreeks toegang op de databases mogelijk zijn via specifieke invoertools. Daarvoor worden diverse (op de markt verkrijgbare) instrumenten ingezet. Hetzelfde geldt ook voor de geo-informatie, waarvoor GIS-systemen worden gebruikt.



Figuur 1. Globale applicatiearchitectuur voor de geo-informatie infrastructuur.

Huidige applicatiearchitectuur MII

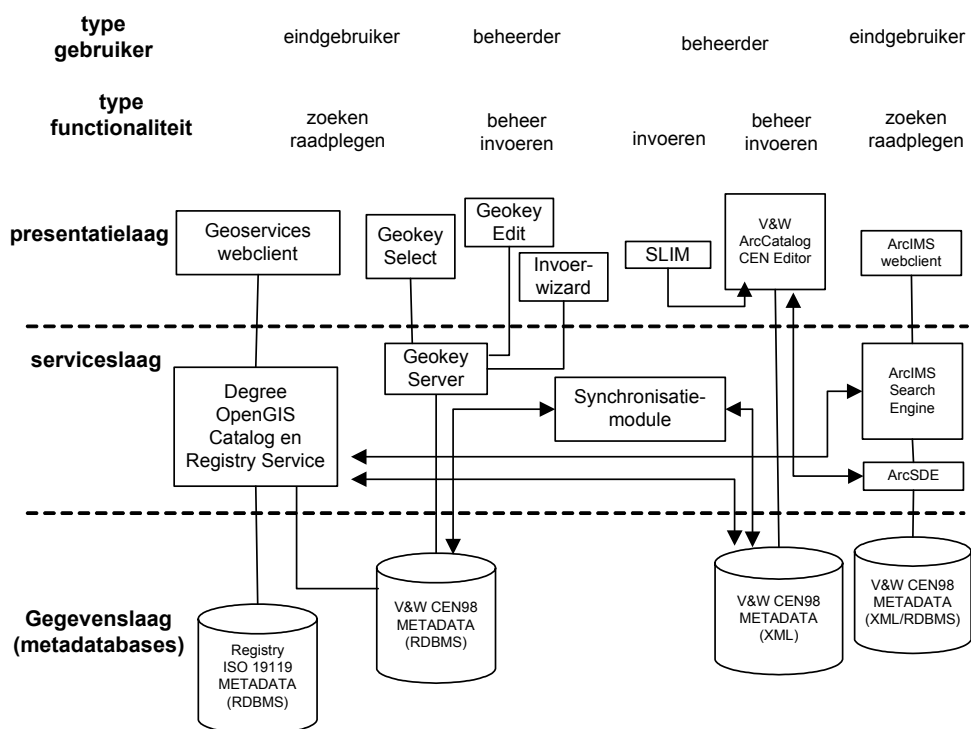
In de huidige situatie heeft de ontwikkeling van specifieke MII-tools in de afgelopen jaren geleid tot een applicatiearchitectuur, waarbij twee software productlijnen, te weten GeoKey en ArcCatalog, naast elkaar bestaan en gegevens uitwisselen via een synchronisatiemodule. Onlangs is de OpenGIS web services architectuur geadopteerd en zijn de OpenGIS Web Catalog en Registry Service opgenomen in de MII-architectuur (zie figuur 2 voor de huidige invulling van de applicatiearchitectuur MII binnen Rijkswaterstaat). Tevens is globaal aangegeven welke gebruiker welke instrumenten voor welke taken inzet.

Op dit moment zijn een vijftal instrumenten in gebruik binnen Rijkswaterstaat:

1. GeoKey Select en Edit (GeoKey Invoerwizard);
2. ArcCatalog VenW CEN Editor en ArcIMS (search engine);
3. de Geokey-ArcCatalog synchronisatiemodule;
4. SLIM: Snel Lokaal Invoeren Metadata;
5. OpenGIS Degree Catalog en Registry service.

Deze instrumenten kunnen afzonderlijk dan wel in combinatie met elkaar ingezet worden. Deze instrumenten be- en verwerken metagegevens conform de richtlijnen van VenW CEN89 [6], ISO 19115 en ISO 19119. In tabel 1 zijn de belangrijkste kenmerken van de bij RWS in gebruik zijnde instrumenten en hun eigenschappen voor de meta-informatievoorziening weergegeven. Vanuit functioneel oogpunt is er enige overlap ontstaan tussen enkele tools. GeoKey Edit en de

ArcCatalog VenW CEN Editor kunnen beide ingezet worden voor het beheer van metadata. SLIM en de Invoerwizard zijn beide instrumenten om op eenvoudige wijze een beperkte set van metadata in te voeren. De met SLIM of GeoKey Invoerwizard ingevoerde metadata dient vervolgens in de metadatabase te plaatsen eerst nog aangevuld te worden conform VenW CEN98 norm. SLIM is afgestemd op de ArcCatalog VenW CEN Editor en de Invoerwizard op GeoKey. GeoKey Select, ArcIMS (search engine) en de OpenGIS Degree Web Catalog en Registry Service verzorgen de functionaliteit voor het zoeken en raadplegen van meta-informatie voor de grote groep eindgebruikers.



Figuur 2. Huidige applicatie-infrastructuur voor de meta-informatievoorziening.

	GeoKey Edit	GeoKey Select	Invoer-Wizard (GeoKey)	SLIM	ArcCatalog VenW CEN Editor	ArcIMS Search Engine	OpenGIS Degree Catalog service	OpenGIS Degree Registry service
Invoeren	v	v	v	v	v			v
Wijzigen	v	v		v	v			v
Zoeken & Raadplegen		v				v	v	v
Datamodel	CEN	CEN	CEN	CEN	CEN	CEN	ISO 19115	ISO 19119
CEN velden	alle	n.v.t.	alle/deel	deel	alle	alle	alle	alle
Standalone	v			v	v			
Netwerk		v	v	(v)	v	v	v	v
Webbased		v	v			v	v	v
Database	RDBMS	RDBMS	Voorportaal RDBMS	XML	XML	DBMS	XML, DBMS	DBMS
Leverancier	Geodan	Geodan	Geodan	ESRI	ESRI	ESRI	Latlon	Latlon
Eigenaar	Geodan	Geodan	VenW	VenW	VenW	ESRI	Open Source	Open Source

Tabel 1. Overzicht kenmerken metadatatools.

In het verleden zijn de tools door verschillende diensten van RWS ontwikkeld. De diensten hadden daarin een redelijk grote autonomie. Dientengevolge is de afstemming tussen de gemaakte producten niet altijd optimaal geweest en is een relatief complexe applicatiearchitectuur ontstaan. Vandaar dat het meta-informatiebeleid zich ook toespitst op het uniformeren en standaardiseren van informatiemodellen en applicaties. Hieronder worden de instrumenten en hun implementatie kort beschouwd.

GeoKey productlijn

GeoKey productlijn bestaat uit drie modules: GeoKey Select, GeoKey Edit en de GeoKey extensie voor Arcview. GeoKey Select is de zoekmodule: hiermee kan via de meta-informatie naar (geografische) bestanden gezocht worden en kunnen deze bekeken worden door middel van een viewer. Met GeoKey Edit wordt de metadata ingevoerd, gewijzigd en beheerd. De metadata wordt opgeslagen in een relationele database.

Voor invoer conform de Richtlijnen VenW is een GeoKey Invoerwizard ontwikkeld: een hulpmiddel om het invoeren van metadata te vereenvoudigen (alleen voor ESRI-shape-bestanden).

ArcCatalog VenW CEN Editor

Met de introductie van ArcGIS heeft ESRI ook het metadatasysteem ArcCatalog op de markt gebracht. ArcCatalog bevat onder andere functies voor het bladeren en zoeken naar geografische informatie, opslaan en bekijken van metadata. In opdracht van Rijkswaterstaat heeft ESRI Nederland de ArcCatalog VenW CEN Editor ontwikkeld. Met deze Editor is het mogelijk op eenvoudige wijze metadata in te voeren en te beheren conform VenW CEN98. Metadata in ArcCatalog is te splitsen in twee groepen, 'properties' (of: inherente metadata) en 'documentatie'. Properties zijn direct uit de dataset af te leiden kenmerken en deze worden door ArcCatalog automatisch geogost. Documentatie wordt door de gebruiker ingevuld, met behulp van een aantal invulschermen in de Editor. Metadata in ArcCatalog wordt opgeslagen in XML-file en kan worden bekeken via stylesheets. Een stylesheet definieert welke metadata velden uit de XML-metadata records worden getoond. Voor RWS is een standaard stylesheet op Internet beschikbaar, zodat alle directies en diensten, maar ook aannemers en ingenieursbureau's deze standaard stylesheet kunnen gebruiken. Ook wordt door enkele organisaties binnen Rijkswaterstaat momenteel gebruik gemaakt van ArcIMS (search engine), waarmee het mogelijk is met en via meta-informatie te zoeken. Daarvoor worden de meta-informatiebestanden (in XML) met en via ESRI's ArcSDE in een RDBMS opgeslagen. De ArcIMS metadata search engine is OpenGIS-compliant, waardoor koppeling met de Geoservices OpenGIS Web Catalog Services mogelijk is (zie ook verderop).

Synchronisatiemodule GeoKey-ArcCatalog

Omdat binnen Rijkswaterstaat zowel de GeoKey als ArcCatalog productlijn in gebruik is, is behoefte ontstaan om metadata van beide systemen uit te wisselen. Die koppeling is gerealiseerd in de vorm van een synchronisatiemodule. Deze module wisselt de metadata in een GeoKey metadatabase uit met de in XML-files opgeslagen metadata van ArcCatalog. De synchronisatiemodule is gerealiseerd vanuit een situatie waarin GeoKey en ArcCatalog op basis van hun onderscheiden sterke punten complementair samenwerk(t)en; ArcCatalog heeft als sterk punt dat metadata onlosmakelijk beheerd wordt bij de data en GeoKey biedt gestructureerde zoekfuncties in een gedistribueerde omgeving. De synchronisatiemodule heeft als primair oogmerk om de metadata in de doorzoekbare database van GeoKey te laden en daarmee voor alle gebruikers toegankelijk te

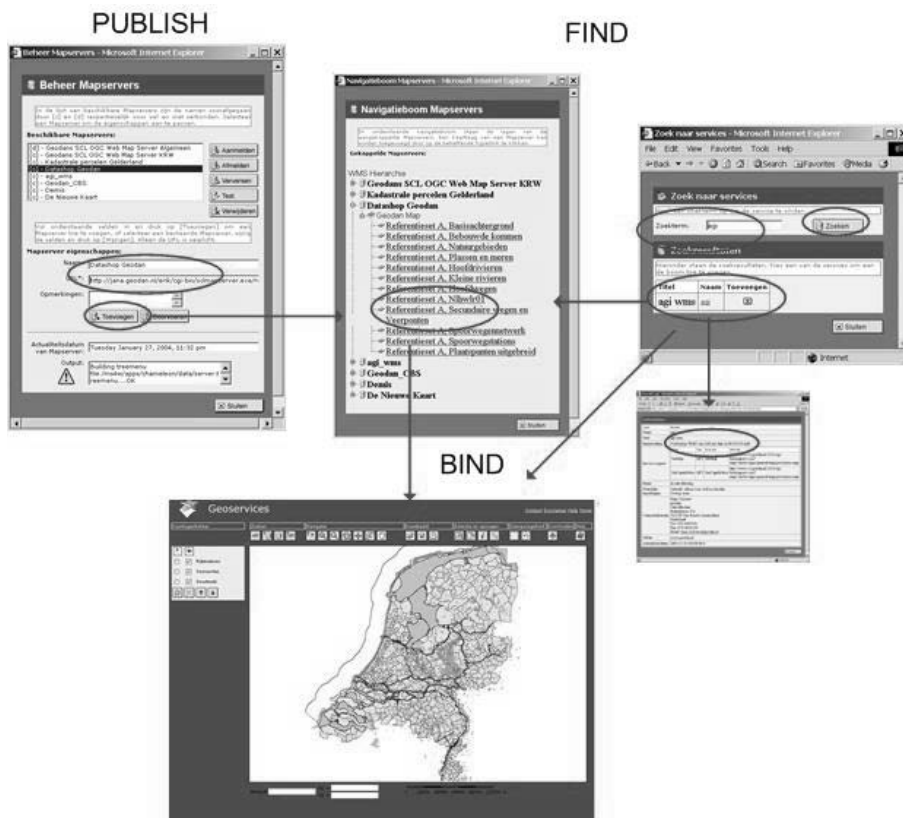
maken. Omgekeerd kunnen metadata records in ArcCatalog vanuit een bestaande GeoKey database worden gevuld.

SLIM: Snel Lokaal Invoeren Metadata

Voor organisatie-onderdelen, die incidenteel relatief eenvoudige bestanden produceren is het toepassen van pakketten als GeoKey of CEN-editor voor het beschrijven van metadata een te zware belasting. Daarom is SLIM ontwikkeld. SLIM staat voor Snel Lokaal Invoeren Metadata. SLIM is een stand-alone applicatie en is niet 'gekoppeld' aan een GIS-applicatie. Dit betekent dat geen inherente meta-informatie ('properties') gegenereerd wordt. De met SLIM gegenereerde metadata (XML) is uitwisselbaar met ArcCatalog en GeoKey. In principe wordt hierin voorzien door het vastleggen van de metadata in XML-tags, die conform de tags in ArcCatalog en GeoKey zijn. SLIM is conform VenW CEN98. SLIM wordt als freeware beschikbaar gesteld.

OpenGIS Web Catalog en Registry Service

Voor het gebruik van de meta-informatie door eindgebruikers - het zoeken, vinden en binden van geo-informatie - is in 2003 binnen Rijkswaterstaat gestart met de implementatie van een centrale OpenGIS web services architectuur. Middels OpenGIS Web Catalog en Registry Services kan een eindgebruiker via de meta-informatie naar de geo-informatie (geografische bestanden en kaarten) en geo-informatiebronnen zoeken en kunnen deze direct opgevraagd en bekeken worden door middel van een web mapping applicatie. Het belangrijkste functionele onderdeel van de OpenGIS achitectuur is de toepassing van het web services principe 'publish-find-bind'. Dit krachtige principe is bij Rijkswaterstaat als volgt geïmplementeerd (zie figuur 3).

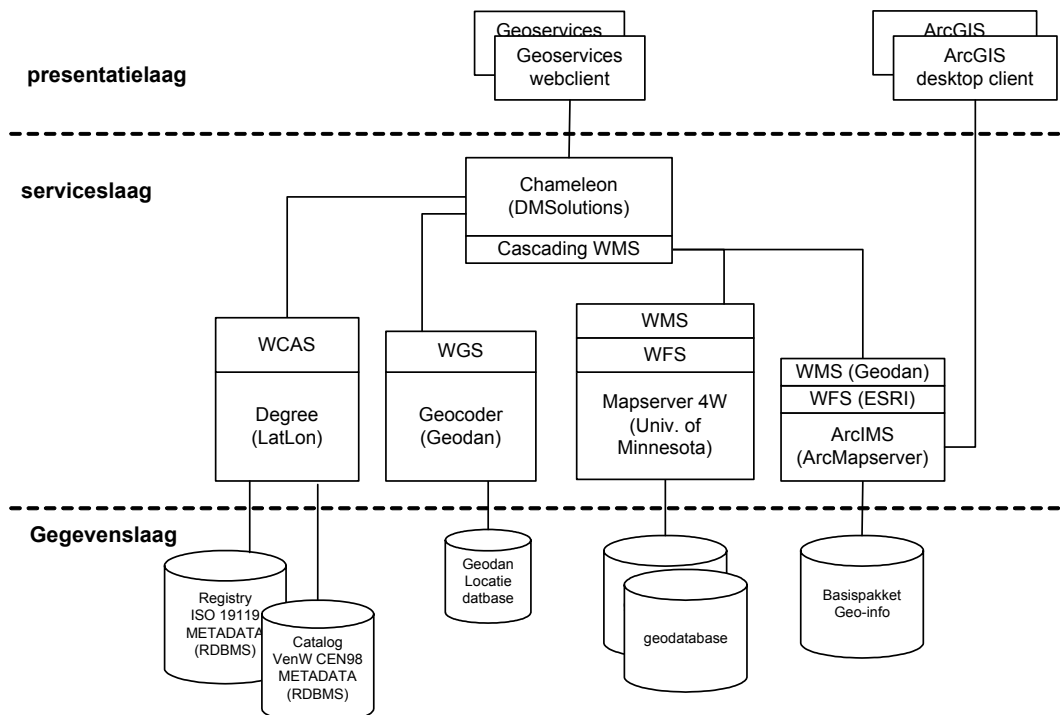


Figuur 3. Het 'publiceren-vinden-binden' principe.

De databeheerder publiceert de meta-informatie over de geo-informatiebron (een kaartenbak) in het register conform ISO 19119. De gebruiker heeft via de applicatie een zoekfunctie ter beschikking waarmee hij of zij in het register naar de geo-informatiebronnen (zgn. 'mapservers') kan zoeken. Een gevonden geo-informatiebron heeft een set van kaarten beschikbaar, die de gebruiker via een navigatieboom kan inzien. Daarna kan de gebruiker de geo-informatie (kaarten) ophalen via de web mapping applicatie. Dit principe van publiceren-vinden-binden is gebaseerd op het (gedistribueerd) aanbieden van geo-informatiebronnen en meta-informatie en hun behorende ontsluitingsdiensten over het internet. Het delen van de geo-informatie staat daarbij voorop en de OpenGIS web services architectuur stelt de gebruikers in staat om zonder hindernissen geo-informatie te zoeken en die meteen te gebruiken. Daarnaast kunnen gebruikers direct zoeken naar de beschikbare geo-informatie (kaarten) conform VenW CEN98 (of ISO 19115).

De binnen Rijkswaterstaat geïmplementeerde OpenGIS software-architectuur kenmerkt zich door een drie-lagen architectuur, waarin presentatie-, service- en gegevenslaag onderling gescheiden zijn (zie figuur 4). Hierdoor is de InternetGIS infrastructuur modulair van opzet en goed schaalbaar. De kern van de architectuur wordt gevormd door de serviceslaag. De modulaire opzet komt in de serviceslaag tot uiting door het gebruik van verschillende software services. Voor de OpenGIS Web Catalog Service (WCAS) wordt gebruik gemaakt van product Degree van het Duitse bedrijf LatLon. Dit product levert zowel de Catalog Service als de Registry Service, zodat het zoeken op zowel geo-informatie als de geo-informatiebronnen mogelijk is. De Catalog Service (ISO 19115 compliant) werkt op een VenW CEN98 database (RDBMS) en voert daarbij een mapping uit van VenW CEN98 naar ISO 19115. De Registry Service werkt op een ISO 19119 database.

Voor haar OpenGIS Web Map Server (WMS) en Web Feature Server (WFS) heeft Rijkswaterstaat gekozen voor de Minnesota MapServer (MS4W) van de Universiteit van Minnesota en

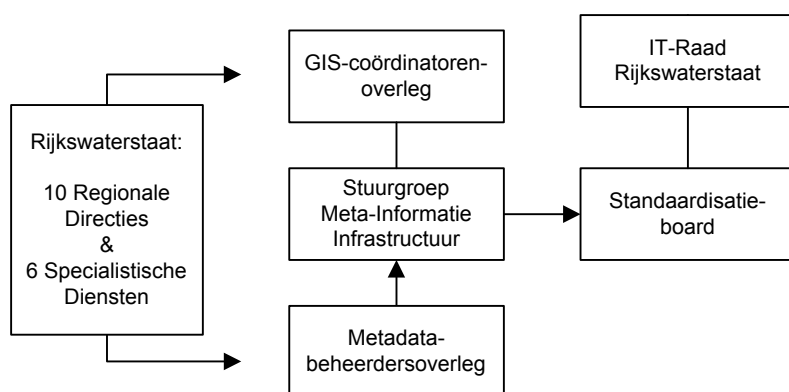


Figuur 4. Catalog service in de OpenGIS web services architectuur van Rijkswaterstaat.

ESRI's ArcIMS uitgerust met een OpenGIS WMS en WFS connector. ArcIMS wordt tevens ingezet om kaarten via webtechnologie naar de desktop ArcGIS client te serveren. Een vrij specifieke servicecomponent is de Web Gazetteer Service (WGS). Deze service (geleverd door Geodan) stelt gebruikers in staat om via een adressen-, postcode- of plaatsnamenregister snel naar een locatie te zoomen. De clientsoftware wordt gevormd door Chameleon, een server-side scripting taal. Chameleon is een product van het Canadese DMSolutions en zorgt ervoor dat de services op de juiste wijze worden aangesproken en resultaten op de juiste wijze bij de gebruiker terechtkomen.

Organisatie en kennisborging

De realisatie van het meta-informatiebeleid en geformuleerde missie is slechts mogelijk als deze blijvend door de gehele organisatie worden ondersteund: "Het beheer en onderhoud van de meta-informatie infrastructuur is organisatorisch ingebed om een duurzame geo-informatievoorziening te kunnen garanderen. De organisatie, kennis en financiën nodig voor de realisatie van de meta-informatie infrastructuur zijn optimaal georganiseerd" [4]. Dat kan het beste door de coördinatie van de meta-informatievoorziening een zichtbare plek binnen de organisatie te geven. Voor de ontwikkeling en implementatie van de Meta-Informatie Infrastructuur (MII) voor geo-informatie is begin 2001 een Stuurgroep MII gevormd die verantwoordelijk is voor het beheren en verder ontwikkelen van de MII (zie figuur 5).



Figuur 5. Organisatie MII bij Rijkswaterstaat.

De stuurgroep is voornamelijk samengesteld uit enkele leden van het GIS-coördinatorenoverleg van Rijkswaterstaat, waarin vertegenwoordigers van de Regionale Directies en Specialistische Diensten zitting hebben. De stuurgroep stuurt enkele werkgroepen aan, die projecten uitvoeren, zoals het ontwikkelen en beheren van tools en richtlijnen omtrent het gebruik van meta-informatie. Daarvoor wordt jaarlijks een werkprogramma opgesteld. De Adviesdienst Geo-Informatie en ICT (AGI) faciliteert de stuurgroep bij het uitvoeren van de projecten. Voor uniformering en standaardisatie wordt een beroep gedaan op de Standaardisatieboard. Uiteindelijk stelt de IT-raad de standaarden vast. Inmiddels is VenW CEN98 als standaard door de IT-Raad vastgesteld en is de OpenGIS web services architectuur als standaard in voorbereiding. Voor meer operationele afstemming vindt een metadatabeheerdersoverleg plaats. Het betreft vooral het werken met de tools en het toepassen van de beschikbare richtlijnen. Ook worden kennis en ervaring onderling uitgewisseld en vindt - indien nodig - advisering plaats aan de stuurgroep. Vanuit deze organisatie vindt waarborging plaats voor het instandhouden van een efficiënte en effectieve MII, zoals het waarborgen van kennis en het beschikbaar stellen van financiën. Kennis wordt via bovenstaande organisatie zoveel mogelijk gewaarborgd. Gewaarborgd moet zijn

dat voldoende financiën beschikbaar zijn om de benodigde investeringen voor implementatie en kennisontwikkeling te kunnen doen. De investeringen in meta-informatie zijn in eerste instantie van organisatorische en personele aard. Daarnaast is de technische infrastructuur in de vorm van de ontwikkeling, het onderhoud en beheer van de MII-tools een structurele financiële post.

Conclusies

Alle 10 Regionale Directies en 6 Specialistische Diensten binnen Rijkswaterstaat beschikken over instrumenten en een organisatie om de essentiële meta-informatie in de geo-informatievoorziening vast te leggen. In de afgelopen jaren is daar voortvarend aan gewerkt. De organisatiebrede 'ontsluiting' van geodatabases en de bijbehorende meta-informatie is noodzakelijk voor het kunnen realiseren van de geformuleerde missie: "In 2007 heeft iedere VenW medewerker ongeacht zijn locatie - thuis, op kantoor en in het veld - de beschikking over een intelligent zoekstelsel om 80% van alle binnen VenW aanwezige geo-informatie te vinden en op te vragen". Vanwege de diversiteit aan systemen bleek het lastig de decentraal opgeslagen meta-informatie organisatiebreed te ontsluiten. De ontwikkelingen rondom OpenGIS Web Catalog Services bieden veel perspectief om de ontsluiting van geo-informatie via het publiceren-vinden-binden principe te realiseren. Met name in het verbeteren van de ontsluiting van geo-informatie speelt de meta-informatie een cruciale rol; meta-informatie is essentieel voor het zoeken, vinden en binden van de geo-informatie. Want meta-informatie is en blijft de sleutel tot de vele geo-informatiebronnen, die nodig zijn voor het uitvoeren van het dagelijks werk in vele organisaties. Het (uit)voeren van meta-informatiebeleid is een belangrijke randvoorwaarde om de meta-informatievoorziening die 'sleutelpositie' in te laten nemen in de geo-informatie infrastructuur.

Referenties

- [1] Grothe, M. en H.J. Scholten (1996), GIS in de publieke sector, een inventarisatie naar gebruik van geo-informatie en GIS bij de Nederlandse overheid. Utrecht: Koninklijk Nederlands Aardrijkskundig Genootschap/Vakgroep Ruimtelijke Economie Vrije Universiteit Amsterdam (Nederlandse Geografische Studies 204).
- [2] Lieshout, M. van en H. Puylaart (2002), Geo-ICT - Trendanalyse. TNO-rapport STB-02-49.
- [3] Grothe, M. (2003) Geo-ICT op koers! Ontsluiting van geo-informatie blijft de grootste uitdaging!, GeoINfo 2003-0, pp. 42-46.
- [4] Visie MII V&W 2002-2007, Ten behoeve van realisatie van een V&W-brede Meta-Informatie Infrastructuur (MII) voor de geo-informatievoorziening, september 2003, AGI-GAG-2003-36.
- [5] Handreiking voor het Programma van Eisen voor de levering van Geografische Informatie bij verkenningen en planstudies van infrastructurele projecten, 8 september 2003, Bouwdienst.
- [6] Richtlijnen voor de implementatie van de Nederlandse voornorm NVN-ENV 12657, versie, 3.0 AGI-GAG-2004-7, Adviesdienst Geo-informatie en ICT.
- [7] Aalders, H.G.J.L en M. Reuvers (2004), 2004, het begin van een nieuw normentijdperk. GeoINfo 2004-1, pp. 20-27.
- [8] Open GIS Consortium (2003), Open GIS Reference Guide, OGC 03-04.

Informatiefabrieken en meta-informatie

Ir. Hein J.M. Corstens
URBIDATA bv
hein.corstens@urbidata.nl

Samenvatting

Informatisering houdt in dat fysieke productieprocessen getransformeerd worden tot informatieprocessen, processen van informatieproductie en -beheer. Deze dienen doelmatig te worden opgezet, als een soort fabriek, een informatiefabriek dus. Daarin wordt informatie geproduceerd uit gegevens, zoals die voorkomen in de veelheid en diversiteit aan bestaande applicaties en zoals die uit het internetverkeer onttrokken worden. Deze gegevens worden geïntegreerd, hetzij voor opslag in datawarehouses en operational data stores, waarna ze omgezet worden in informatie voor ondersteuning van operationele, tactische of strategische beslissingen, hetzij voor direct gebruik in applicaties (waarmee Enterprise Application Integration (EAI) vorm krijgt). Voor die integratie heeft URBIDATA een hulpmiddel ontwikkeld, de Universal Data Store (UDS). De informatiefabriek dient bestuurd te worden, net zoals een gewone fabriek. Daarvoor is informatie nodig, in dit geval meta-informatie. Meta-informatie – die gemaakt wordt uit metadata – is daarmee de onmisbare basis voor informatisering. De informatisering wordt al snel gevolgd door een volgende fase, die van de meta-informatisering. Daarin dient integratie tot stand te komen van metadata uit verschillende bronnen, gericht op de besturing van de 'meta-informatiefabriek'.

Besturingsmodel

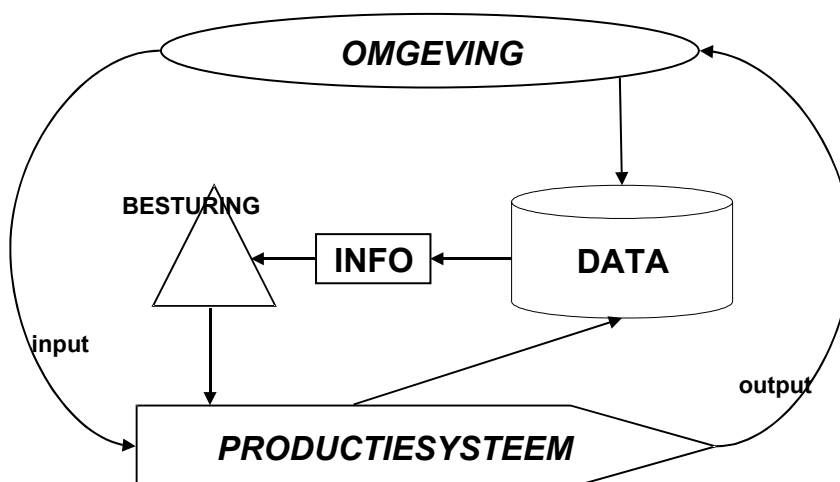
De informatie-infrastructuur wordt steeds belangrijker voor bedrijven en overheidsorganisaties. De informatie-infrastructuur is het geheel van voorzieningen voor algemeen gebruik door en toegankelijk voor alle bij het primaire proces betrokken organisatieonderdelen. De informatie-infrastructuur bestaat uit componenten voor ontwikkeling en beheer van gegevens, applicaties, configuraties, communicatie en organisatie. De informatie-infrastructuur beschouwen wij als een geheel dat BESTUURD moet worden en wel zodanig dat een optimale aanpassing tussen informatie-infrastructuur en haar omgeving gerealiseerd wordt. In dat opzicht is een informatie-infrastructuur wel te vergelijken met een ecosysteem. Net zoals ecosystemen in de natuur moet een informatie-ecosysteem aanpasbaar zijn. Op den duur veranderen zowel het evenwicht tussen verschillende componenten en hun onderlinge relaties als de omgeving. Een gezond ecosysteem ontstaat alleen als de componenten aanpasbaar, veranderbaar en in evenwicht zijn. Nu moeten die kenmerken uiteraard wel gericht zijn op een bepaald doel, bijvoorbeeld continuïteit. De onderlinge aanpassing van systeem en omgeving is daarom te zien als een besturingsproces. Daarin speelt terugkoppeling een essentiële rol. In plaats van informatiesystemen en informatie-infrastructuren te vergelijken met gebouwen en de informaticus met een architect trek ik liever een vergelijking met een landschap of desnoods een stad, welke voortdurend groeit en bloeit en waar men doelen tracht te bereiken door voortdurend kleine ingrepen te treffen, waarna men het systeem zijn gang laat gaan. Voortdurend wordt er geëvalueerd, gevolgd door bijsturing. Terzijde merk ik op dat ik in dat verband een vraagteken plaats bij de manier waarop sommigen denken de plannen van de Nederlandse overheid voor authentieke registraties te realiseren, namelijk als een geheel dat volgens een blauwdruk 'weggezet' moet worden.

Maar ter zake. Wat is besturing? Besturing is het voortdurend corrigeren van een systeem in de richting van een doel. Letterlijk zien we dat in de besturing van een auto. Afhankelijk van de termijn waarop het doel bereikt moet worden spreken we over strategische, tactische dan wel operationele besturing. Voorbeeld: bij het geven van een inleiding is het strategische doel de aanwezigen te informeren. Het tactische doel is de aanwezigen te vermaken. Als de inleider ziet dat er meerdere aanwezigen in slaap vallen moet er bijgestuurd worden, bijvoorbeeld door het luid vertellen van een leuke anekdote. Als nu de aandacht teveel naar bijzaken gaat moet er weer bijgestuurd worden richting onderwerp. De operationele besturing betreft onder meer de bediening van de laptop en de projector.

De toepassing van genoemde driedeling op een bedrijf levert de volgende indeling in informatiesystemen op:

- business intelligence systemen voor de strategische besturing (strategische planning; research, analyse en voorspelling);
- business management systemen voor de tactische besturing;
- business operations systemen voor de operationele besturing (dit zijn veelal legacy systemen).

Het besturingproces wordt gerepresenteerd door onderstaand plaatje.



Besturingsmodel.

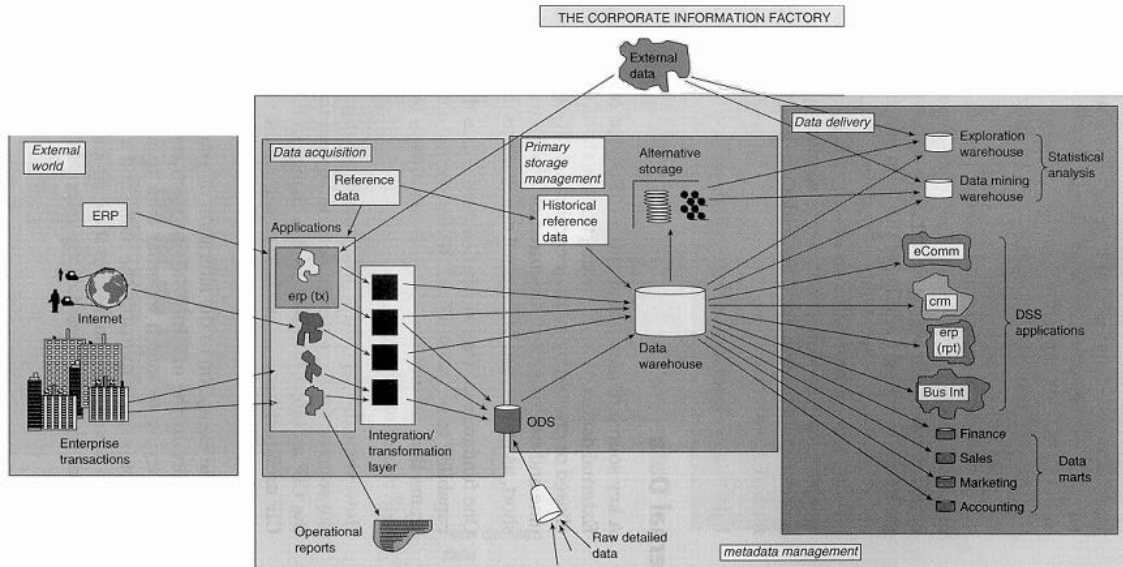
De besturing heeft betrekking op een productiesysteem, dat via input en output in relatie staat met de omgeving. De besturing – symbolisch als een piramide weergegeven met de strategische besturing aan de top en de operationele besturing aan de basis – vindt plaats op basis van informatie, die gemaakt wordt uit data, die betrokken worden uit het productiesysteem zelf en de omgeving.

Informatisering en informatiefabrieken

Informatisering is een ontwikkeling, waarbij fysieke productieprocessen langzaam maar zeker getransformeerd worden tot informatieprocessen. Daarin bestaat de dagelijkse activiteit uit verwerking van gegevens over de productie zelf én over de omgeving waarin dit plaatsvindt tot informatie, waarmee het productieproces wordt bestuurd. Het productieproces zelf is geautomatiseerd. Een fase verder is de besturing van de informatieprocessen. Daarbij worden gegevens

over het informatieproces verwerkt tot informatie op een hoger niveau. Het informatieproces zelf wordt daarbij geautomatiseerd. Uiteindelijk ontstaat een 'informatiefabriek', een geheel van processen waarin de grondstof gegevens verwerkt wordt tot informatie.

Door Bill Inmon, één van de grondleggers van Data Warehousing, is het concept Corporate Information Factory (CIF) uitgewerkt [Inmon e.a.]. In onderstaande figuur wordt de CIF weergegeven.



Corporate Information Factory (bron: [Inmon e.a.]).

Een CIF lijkt veel op een gewone fabriek *. Grondstoffen en onderdelen gaan een fabriek in en worden opgeslagen. De ruwe goederen worden in assembleerlijnen omgevormd tot een diversiteit van halffabrikaten en eindproducten. De samenstellende delen van de CIF zijn:

- Buitenwereld: dit is het geheel van processen, waarvoor de informatie uiteindelijk bedoeld is en van waaruit de gegevens het systeem binnenkomen.
- Applicaties deze zorgen voor verzameling van gedetailleerde transactiegegevens, directe interactie met de gebruiker, controle en aanpassing van gegevens en redactie van data. Veelal vormen de applicaties een onsamenhangend geheel met data die niet geïntegreerd zijn, het is 'legacy'. De nadruk ligt op snelle response. In toenemende mate is het internet de bron van de data die tot grondstof van de informatiefabriek dienen.
- Integratie- en transformatielaag: deze bestaat uit programma's die de niet geïntegreerde data uit de applicaties combineren en transformeren tot corporate data.
- Operational Data Store (ODS): dit is een hybride structuur, die enerzijds op operationele processen, anderzijds op besluitvormingsondersteuning is gericht. Deze structuur maakt de ODS de meest complexe component van de CIF. De data in de ODS zijn geïntegreerd, vluchtig (ze worden geactualiseerd als onderdeel van operationele processen), actueel en gedetailleerd.
- Data Warehouse, de basis voor alle strategische beslissingsondersteunende processen. De data in het data warehouse zijn: 'subject oriented' (gericht op onderwerpen zoals Klanten,

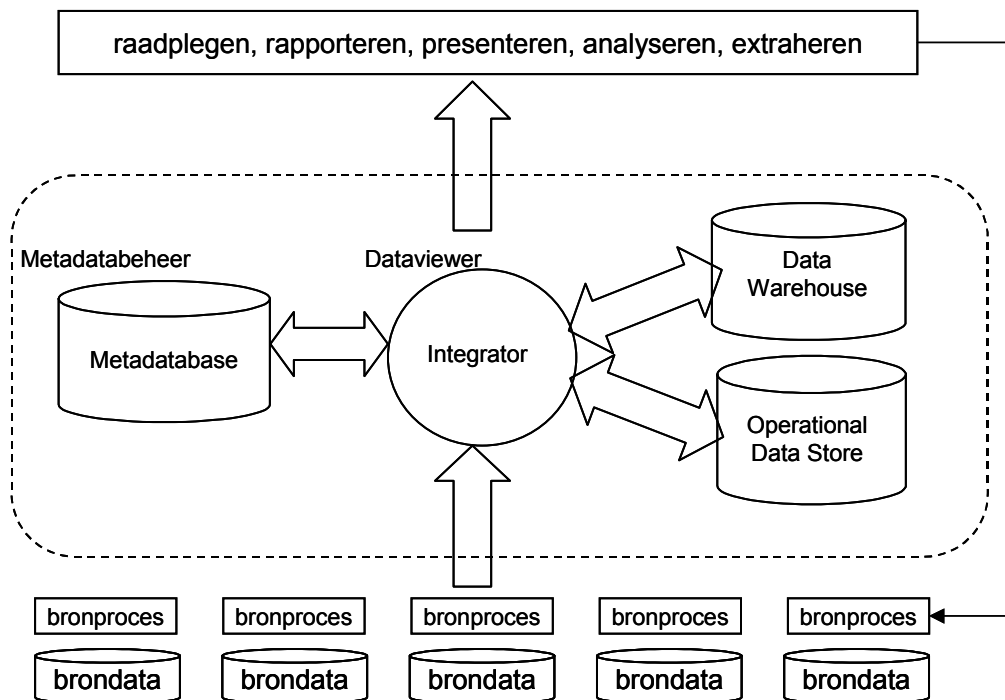
* Inmiddels hebben Inmon en de zijnen een variant van de CIF voor de overheid bedacht: de GIF, wat staat voor 'Governmental Information Factory'.

Producten, Verkopers, Transacties, Orders, Policies, Accounts, Verplaatsingen), dus NIET functie- of applicatiegericht, geïntegreerd en tijdsvariant (ieder record heeft een geldigheidsperiode of -moment). Ze worden eenmalig opgeslagen en blijven onveranderd in het Data Warehouse staan. Ze zijn niet-vluchtig (er wordt gewerkt met snapshots) en zowel samenvattend als gedetailleerd. De datastructuur is in principe genormaliseerd en wel om redundantie te elimineren.

- Data Marts: een mart is een collectie data, toegesneden op de besluitvormingsondersteuning van een bepaalde afdeling. Een data mart is meestal gedenormaliseerd. Data Marts worden in het leven geroepen ten behoeve van beheersbaarheid, kosten en aanpasbaarheid aan wensen van de gebruiker. Data Marts worden vanuit het Data Warehouse gevuld.
- Exploration and Data Mining Warehouses: deze componenten worden gecreëerd in verband met de zeer grote queries die alle capaciteit verbruiken. Het Exploration Warehouse is een fysiek afgescheiden structuur die exclusief gericht is op onderzoeksprocessen. Het is breed opgezet en bevat veel soorten data en relaties tussen de data; doel is hier het onderkennen van patronen en het formuleren van hypothesen. Daarna ontstaat er behoefte aan een Data Mining Warehouse. Dat bevat veel feitelijke gegevens met weinig variatie in datatype; de data miner is meer gericht op het testen en valideren van hypothesen.
- Alternative Storage Component voor de tijdelijke of permanente archivering van de data.
- Internet/Intranet Components. In toenemende mate vinden interactie en communicatie van de CIF met de buitenwereld plaats via Internet. Intern geschiedt dit via Intranet.
- Metadata; hierop komen we later terug.

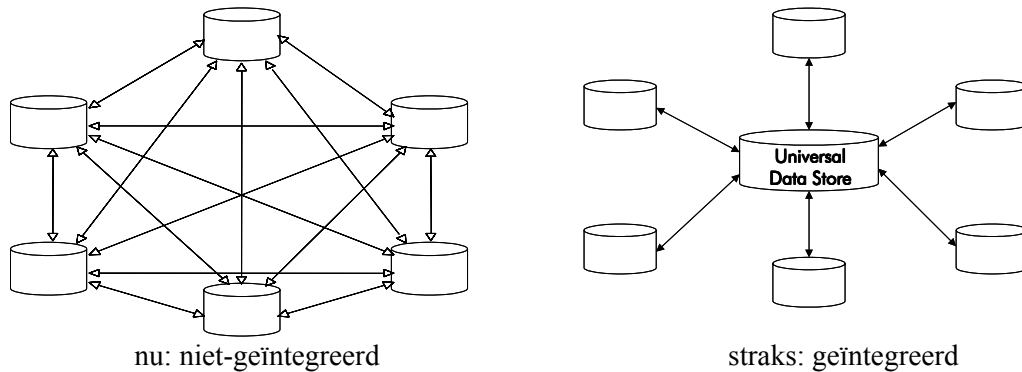
Universal Data Store

URBIDATA heeft als basis voor informatiefabrieken integratiesoftware ontwikkeld, UDS – Universal Data Store – genaamd.



Universal Data Store.

De UDS is specifiek gericht op zowel RUIMTELIJKE als ADMINISTRATIEVE DATA en op TERUGKOPPELING naar de bronsystemen. Daardoor kan een vereenvoudiging in de informatie-infrastructuur bewerkstelligd worden, zoals de volgende figuur illustreert.

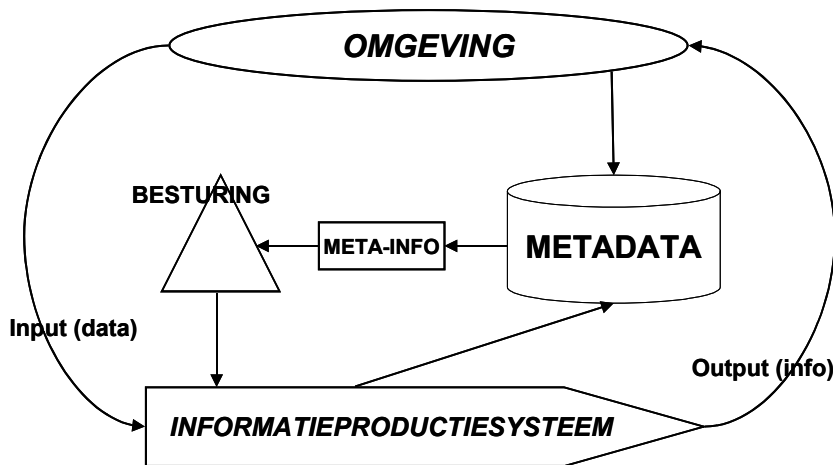


Integratie met behulp van de UDS.

Terzijde: de niet-geïntegreerde situatie wordt wel aangeduid als 'spaghetti'. De geïntegreerde oplossing is 'ravioli', een gestructureerd geheel van objecten, waarin alle gegevens en methoden betreffende één objecttype zijn samengebald. In de wereld van de geografische informatiesystemen wordt wel een tussenoplossing gebruikt, waarbij door de koppeling van kaartlagen op grond van ligging administratieve gegevens onderling gerelateerd worden. Deze handige 20/80 oplossing noemen we 'lasagne'.

Meta-informatie

Ook de informatiefabriek moet bestuurd worden. Dit wordt in onderstaand plaatje weergegeven.



Metabesturingsmodel.

We zien dat er eigenlijk geen verschil bestaat met het besturingsmodel voor een gewone fabriek. We moeten alleen de fysieke objecten vervangen door informatieobjecten.

Meta-informatie wordt gevormd uit metadata, data over data. Deze worden in de UDS opgeslagen in een metadatabase en beheerd in de metadatamanager. In de metadata worden beschreven:

- ontwerpgegevens: datamodellen, functionele ontwerpen, technische ontwerpen, kwaliteitseisen, enz.;
- productiegegevens: fysieke modellen, testen, planning en voortgang;
- gegevens over levering en gebruik: kostenverrekening, productencatalogus, planning en logistiek.

In concreto wordt de kern van de metadata gevormd door:

- gedetailleerde beschrijving van de data (datasets, objecten, attributen, relaties, constraints), zowel qua syntax (vorm) als semantiek (inhoud);
- specificatie van de applicaties waarmee de data beheerd en benaderd worden, alsmede de bijbehorende platforms;
- relatie met de organisatie en de taken die vervuld worden;
- transformatieregels en processen;
- verversingsplanning;
- gebruik;
- verwijzingen naar documentatie (literatuur, websites);
- specifieke kenmerken; bijvoorbeeld specifieke geodetische eigenschappen.

Ook metadata worden in toenemende mate via het internet betrokken. Metadata vormen de lijm die de componenten van de CIF bij elkaar houdt.

Het beheer van metadata heeft zeer veel aandacht, net zo als dat voor andere data geldt. In de UDS is de MetaDataManager dan ook het hart van het systeem. Belangrijk is de afbakening tussen centraal en decentraal beheer van metadata, tussen de behoefte aan autonomie en de behoefte aan samenwerken en elkaar informeren. Er dient daarin een evenwicht bereikt te worden. In iedere component van de CIF dient er een scheiding aangebracht te worden tussen gemeenschappelijke en autonome metadata. Per gemeenschappelijk metadata-element dient vastgesteld te worden welke component dit mag muteren. Het is voor alle andere componenten zichtbaar.

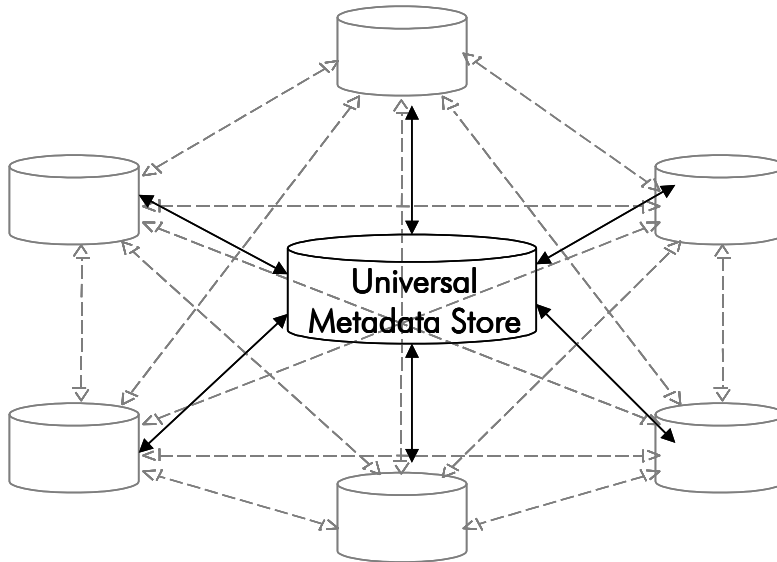
Het verkrijgen van metadata is in het algemeen een groot probleem. Aanbevolen wordt bij de ontwikkeling van de CIF te werken met tools die automatisch metadata produceren.

Overigens wordt de UDS ook ingezet om metadata te integreren. De diverse bronmetadata worden door de UDS geïntegreerd en volgens bepaalde regels aan verschillende gebruikersgroepen ter beschikking gesteld. De UDS werkt in dat verband dus als 'UmetaDS' of 'UmDS'. Dit brengt ons tenslotte op Enterprise Application Integration.

Enterprise Application Integration (EAI)

De UDS kan een belangrijke rol spelen in 'Enterprise Application Integration' (EAI). Dit is het integreren van de verschillende systemen tot virtueel één systeem. Voor een gebruiker maakt het daarbij niet uit of er één groot systeem is, dan wel een verzameling van met elkaar verbonden kleine systemen. Er kan zelfs per gebruiker – door de creatie van 'portals' – een eigen view op het geheel van informatiesystemen gemaakt worden. Met zo'n portal is er integratie op user interface niveau ('sceenscraping'). EAI vergt echter integratie op dataniveau. Daarbij lezen applicaties elkaars gegevens en schrijven er ook naar weg. Dit gebeurt door 'message brokers', die gegevens routeren en transformeren (aangepast aan het datamodel van de doelapplicatie). Door gebruik van de UDS wordt consistentie gegarandeerd door toepassing van de in de metadatabase opgeslagen transformatieregels.

In feite is een centrale data store dan niet altijd meer nodig: in plaats daarvan kan men een centrale metadata store inrichten, waarin alle metadata, transformatieregels en de planning en scheduling daarvan zijn opgenomen. Met behulp daarvan kan op dataniveau onderlinge uitwisseling gestuurd worden.



Structurering door centralisatie van metadata.

Het zal duidelijk zijn dat zo iets ook voor de metadata zelf gerealiseerd zou kunnen worden. Uiteindelijk blijft er dan alleen nog centraal beheer op metaniveau over.

Conclusie

Informatisering, de transformatie van fysieke productieprocessen tot informatieprocessen, krijgt een gestructureerde vorm in informatiefabrieken, waarin informatie wordt geproduceerd uit bestaande gegevens. De informatiefabriek dient bestuurd te worden op basis van meta-informatie. De informatisering wordt al snel gevolgd door een volgende fase, die van de meta-informatisering. Daarin dient integratie tot stand te komen van metadata uit verschillende bronnen, gericht op de besturing van de 'meta-informatiefabriek'.

Literatuur

[Inmon e.a.]

W.H. Inmon, C. Imhoff, R. Sousa, 'Corporate Information Factory', 2001 ISBN: 0-471-39962-2

SuperGlue: De lijm tussen alle metadata

Bert Oosterhof, Director of Technology EMEA
Informatica
boosterhof@informatica.com

Inleiding

Michael Smith, oprichter van het wereldwijd bekende bedrijf Federal Express, zei eens: "De informatie over een pakket is voor ons even waardevol als de aflevering ervan." Vertaald naar onze industrie, de informatie technologie, zouden we kunnen zeggen: "De informatie over de gegevens (de metadata) is net zo belangrijk als de gegevens zelf." Deze laatste stelling wordt echter nog door slechts weinigen onderschreven. De hoeveelheid metadata in organisaties is enorm, de verscheidenheid eveneens. De waarde van die metadata zo mogelijk nog groter, maar de toegang ertoe, het inzicht erin het benaderen ervan vormt vaak een groot probleem. Dit artikel beschrijft de beknopt de diverse uitdagingen en hoe Informatica met het product SuperGlue een hulpmiddel heeft ontwikkeld om deze problemen op te lossen. Er is bewust weinig aandacht besteed aan Geografische Metadata, aangezien andere presentaties daar uitgebreid op ingaan. Het doel van artikel is om belang van bedrijfsbrede metadata aan te geven, en een mogelijke oplossing aan te dragen om alle metadata te beheren en om vanuit een centrale omgeving inzichtelijk te maken.

Soorten Metadata

Wat is meta-data eigenlijk? Een zeer eenvoudige definitie is: metadata is data (gegevens) over de data. Dit helpt ons echter nauwelijks verder. Bovendien leidt deze definitie niet tot eenduidigheid. Het datatype zegt iets over een klant nummer, net als het feit dat klantinformatie in de tabel KLANT is terug te vinden. Het feit dat een bepaalde klant wel of niet zijn rekeningen betaald is ook informatie over de Klant, maar is geen metadata. Vanuit een systeemontwikkelingsperspectief zouden we kunnen zeggen: metadata is alle informatie die we nodig hebben om applicaties te kunnen ontwerpen, bouwen, implementeren en onderhouden. Een operationele definitie zou kunnen zijn: alle informatie over het gebruik van digitale gegevens in een organisatie, terwijl de eindgebruiker of 'business user' alle informatie, definities, betekenissen over gegevens, over informatie, over kennis-bronnen als metadata bestempelen. De eerste conclusie is dan ook dat we zeker drie soorten metadata kunnen onderscheiden: technische, operationele en inhoudelijke metadata.

Het belang van Metadata

Een eenvoudig lijkende vraag als: "Wat zijn onze 10 meest winstgevende klanten?" kan lang niet door alle bedrijven worden beantwoord. En als een management informatie rapport beschikbaar is die deze vraag beantwoordt, dan levert het nieuwe vragen op. Wat betekent winstgevend? Hoe is dat berekend? Uit welke systemen is hiervoor data gebruikt? Over welke periode gaat het? Hou oud is de gebruikte data? Hoe definiëren we het begrip 'klant'? Dit lijken misschien irrelevante vragen, maar als van een onderneming de ene divisie Philips N.V. als klant rekent, en een andere divisie gebruikt een ander systeem, waarbij elke dochter-onderneming (Medical, Consumer Electronics, etc.) als klant wordt gezien, dan heeft dat gevolgen voor het resultaat van de winst per klant. Ook maakt het verschil of bijvoorbeeld de kosten van de help-

desk of call-center per klant wordt doorberekend of dat deze kosten als algemene kosten worden gezien. Kortom, metadata plaatst data in context en maakt data begrijpelijk en dus bruikbaar en waardevol!

Sinds een aantal jaren wordt de regelgeving door overheden en nationale banken voor bedrijven en organisaties aanmerkelijk aangescherpt. Dit heeft soms internationale consequenties. Zo kan een nieuwe wet in Amerika (bijv. Sarbanes-Oxley act, waarbij ondertekenaars van kwartaal- en jaarrapporten persoonlijk aansprakelijk zijn voor de juistheid hiervan) gevolgen voor alle ondernemingen, ook Nederlandse, die in Amerika aan de beurs genoteerd zijn. Deze regelgeving heeft de vraag naar een gedegen metadata management oplossing enorm gestimuleerd, aangezien het belang van inzicht in, betrouwbaarheid en controleerbaarheid van data en rapportages nu op hoog nivo in de organisaties wordt ingezien.

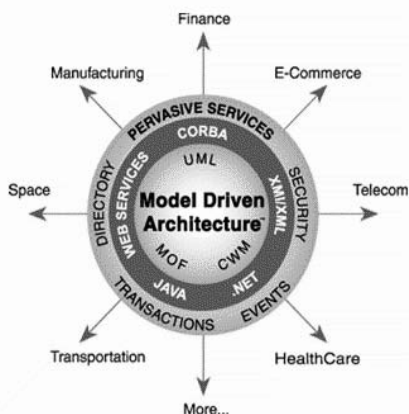
Metadata Bronnen

Organisaties gebruiken voor het automatiseren van hun bedrijfsprocessen talloze producten en/of applicaties. Deze applicaties werken met de voor hen van belang zijnde data en metadata. Hierdoor kan de verscheidenheid aan data en metadata enorm zijn. Data bevinden zich in bestanden. Deze bestanden kunnen beheerd worden door een database management systeem, door een Geografisch Informatie systeem, door een imaging applicatie, door een tekstverwerker, door een design-tool, door een business intelligence product, door een CAD/CAM applicatie, etc. etc. Om een definitie te raadplegen of om berekening te achterhalen moet de betreffende applicatie worden gestart of gebruikt. Simpele, maar toepassing-overschrijdende vragen als: "wat is de impact als we het 'klantnummer' van onze klanten veranderen van 9 naar 10 posities" zijn alleen te beantwoorden door alle gebruikte producten te onderzoeken of 'klantnummer' in die applicatie gebruikt wordt. Hoe kan deze metadata-chaos geordend worden? Het antwoord is: standaardisatie. Aangezien het een utopie is te verwachten dat alle producten en toepassingen binnen afzienbare tijd gebruik zullen maken van dezelfde metadata-standaard, is een goed alternatief: een Metadata Warehouse. Dit concept is door Informatica gebruikt om het product SuperGlue te ontwikkelen.

Informatica SuperGlue

De uitgangspunten

Uitgangspunten bij het ontwikkelen van Superglue zijn geweest: 1. gebaseerd op standaarden, 2. uitbreidbaarheid voor ieder type metadata, 3. open architectuur, 4. bruikbaar voor IT- en Business user en 5. schaalbaarheid.

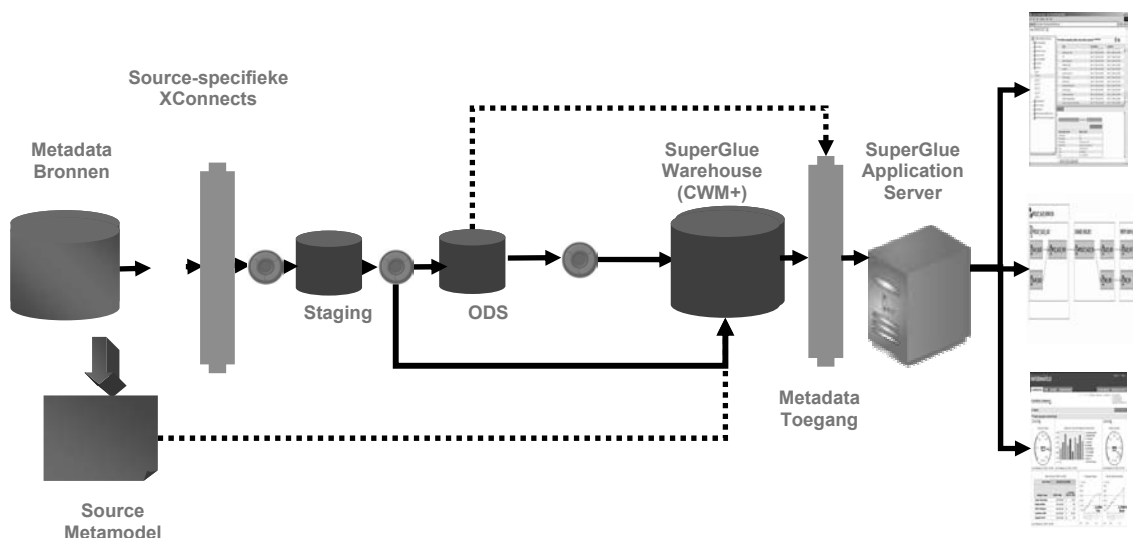


Figuur 1. OMG's Model Driven Architectuur, gebaseerd op metadata.

1. Standaardisatie is op korte termijn moeilijk te realiseren. Dit hebben we gezien aan bijv. De SQL standaard. De meeste DBMS'en zijn hierop gebaseerd, maar een programma dat werkt voor de 1 werkt niet altijd voor de ander. Dit geldt ook voor metadata standaarden. De organisatie die op dit gebied het verst gevorderd is, is de Object management Group (OMG). XMI (metadata uitwisseling), MOF (Metadata Object Facility voor implementatie), CWM (Common Warehouse Metamodel) en MDA (Model Driven Architecture). Hoewel ook andere organisaties zich met standaardisatie bezighouden (zoals Dublin Core), is SuperGlue een implementatie van het CWM dat uitgebreid is op basis van de MOF-standaard en dat metadata uit kan wisselen d.m.v. XMI. Daarnaast is de implementatie ervan gerealiseerd in een relationele database, zodat een bestaand DBMS gebruikt kan worden en zodat met standaard rapportage tools gewerkt kan blijven worden.
2. Door te kiezen voor een implementatie op meta-metadata niveau, kan SuperGlue uitgebreid worden met elke verzameling metadata en toch de semantiek van de bron behouden. De meta-metadata beschrijft het metamodel. Zo heeft een DBMS o.a. tabellen, indexes en triggers, een BI-tool rapporten, metrics, documenten, etc. Door een nieuw metamodel aan SuperGlue toe te voegen, kan daarna de desbetreffende metadata ook worden toegevoegd en herkend.
3. De open architectuur stelt organisaties in staat op bestaande infrastructuur SuperGlue in te zetten. Dit geldt zowel voor het platform (Linux, Unix, NT), het DBMS (DB2, Oracle, SQL Server) en het J2EE platform (WebSphere, WebLogic, Jboss, SunONE).
4. Zoals eerder vermeld is er een nauwe relatie tussen technische, operationele en business-metadata. De toegang tot en inzicht in metadata moet dan ook eenvoudig en begrijpelijk zijn. Vandaar dat gekozen is voor het toegankelijk maken van metadata d.m.v. een web-browser, door integratie in een bedrijfs-portal of via webservices.
5. Als deze metadata benaderd wordt door honderden of zelfs duizenden eindgebruikers, dan moet het platform ook schaalbaar zijn, om acceptabele responsetijden te kunnen garanderen. Om dit te realiseren in een 'open systems' omgeving is gekozen voor een applicatie server platform gebaseerd op J2EE.

De architectuur

SuperGlue is opgebouwd uit een aantal componenten, De basis is de implementatie van OMG's CWM, een standaard Warehouse metamodel. Dit warehouse bevat alle metadata vanuit de ondersteunde metadata bronnen, inclusief de historie ervan, en inclusief de taxonomie van de oor-



Figuur 2. De architectuur van SuperGlue.

spronkelijke bron. Aan de linkerkant van het schema in figuur 2 zijn de zogenaamde Xconnects te zien. Dit zijn specifieke koppelingen om metadata te ontsluiten uit de diverse bronnen. Vervolgens wordt deze metadata middels integratie en transformatie processen in het Warehouse geladen. In deze processen worden ook wijzigingen ontdekt en kunnen versies van metadata objecten worden bijgehouden. Aan de rechterkant van het schema is te zien dat de ontsluiting (rapportages, dashboards, zoek-opdrachten) middels een web-userinterface gerealiseerd is.

De functionaliteit

Gebruikers van SuperGlue hebben een centrale toegang om definities, eigenschappen en andere metadata te vinden. Men kan in alle metadata zoeken, er rapporten over maken en zelfs relaties en afhankelijkheden (lineage) ontdekken. Mits geautoriseerd kunnen gebruikers ook associaties tussen metadata objecten van diverse bronnen aanbrengen. Ook met rapportage tools van andere leveranciers kunnen rapporten en queries gemaakt worden. Tevens wordt voorzien in WebService toegang, zodat bijv. vanuit een applicatie aan Superglue gevraagd kan worden wat de betekenis van een attribuut is, of wie de eigenaar is, of van welke datum de informatie is, etc. Of het nu gaat om technische-, operationele- of business-metadata, SuperGlue biedt de mogelijkheid om deze op een eenvoudige manier terug te vinden.

Samenvatting

Een goed beheer van en inzicht in metadata wordt van steeds groter belang voor grote organisaties. De zgn. point-solutions voorzien in metadata-beheer, dat voor die toepassing van belang is. Aangezien het een utopie is dat alle producten (of leveranciers) hun metadata in één en dezelfde repository zullen opslaan, lijkt een bedrijfsbreed metadata-warehouse een goed alternatief. Door het bieden van een soort meta-data portal, dat uitbreidbaar is, inpasbaar is binnen bestaande infrastructuren en te koppelen aan GIS en andere informatiesystemen, kan Superglue veel bedrijven helpen het metadata beheer te vereenvoudigen.

Metadata en gedistribueerd GIS

Drs. Bert Vermeij
ESRI Nederland
b.vermeij@esrinl.com

Inleiding

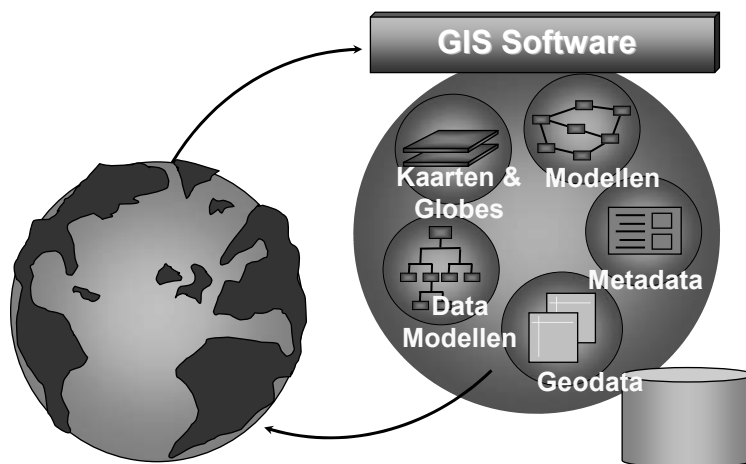
Wat is de overeenkomst tussen een pot pindakaas, een kaart uit de 16^e eeuw, een GIS-bestand en een webservice? Op de pindakaas zit een etiket met merk, gewicht en samenstelling van het product in de pot. De kaart heeft een legenda die vertelt wat de verschillende kleuren en symbolen voorstellen. Bij het GIS-bestand horen technische gegevens als de projectie. Bij webservices hoort een beschrijving van de functies van een service en van de argumenten die moeten worden doorgegeven om deze functies aan te roepen.

Allemaal meta-informatie, die vertelt wat er in een product zit. Meta-informatie is dus niets nieuws en zeker niet uniek voor (geografische) informatievoorziening. Het is de informatie die een producent gebruikt om zijn product te beschrijven en die wij als consument gebruiken om te beoordelen of het product past bij onze behoefte. Zonder metadata geen GIS.

Intelligente GIS systemen

Geografie helpt ons om onze leefomgeving te beheren en te organiseren. GIS-systemen brachten daarbij een enorme stroomversnelling te wege, vooral omdat ze grote datavolumes efficiënt kunnen verwerken. Moderne geografische informatiesystemen abstraheren de werkelijkheid in vijf basiselementen: geodata, kaarten en globes, modellen van werkprocessen, datamodellen en metadata. Gezamenlijk representeren deze vijf componenten onze geografische kennis van de wereld.

Samen met GIS functionaliteit vormen deze vijf basiselementen de bouwstenen voor “intelligente geografische informatiesystemen”.



Intelligente geografische informatiesystemen.

Deze systemen bevatten de kennis over de leefomgeving en zijn daarmee de basis voor beheer en inrichting. Ze stellen ons in staat efficiency van werkprocessen te verbeteren, betere plannen te maken, besluitvorming te ondersteunen en beleid te communiceren.

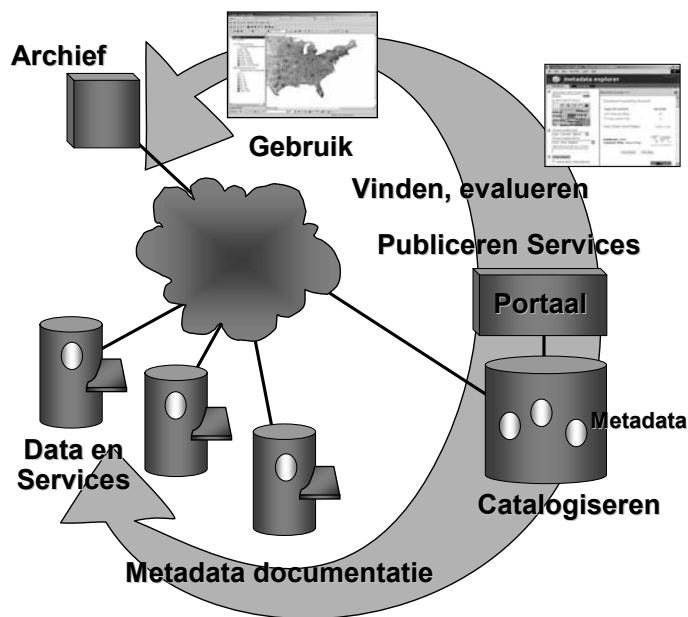
Internet en vooral de nieuwe webservices standaarden voegen een extra dimensie toe. Met webservices kunnen we de kennis in de systemen publiceren en zo delen met anderen. Het internet verbindt de kennis van individuele organisaties tot een geografische kennisinfrastructuur. In deze infrastructuur, die bestaat uit webservices, clients, servers en portalen is meta-informatie de alles verbindende component. Hierover later meer.

De geodatabase bevat dus een abstractie van de wereld in vijf basiselementen, waaronder geïntegreerde metadata. Metadata beschrijft de kaarten, de procesmodellen, de data en de datamodelen en vooral ook de samenhang tussen de individuele componenten.

In de ArcGIS software van ESRI is metadata functionaliteit standaard aanwezig.

Levenscyclus van data

Metadata is onmisbaar geworden. Naarmate organisaties over meer data beschikken wordt het ook steeds belangrijker een goede administratie bij te houden. Met andere woorden, door de gehele levenscyclus van data speelt metadata een rol.



Data levenscyclus.

De levenscyclus begint met het creëren van databestanden (of services). Hier zorgt de bronhouder voor documentatie met metadata. Deze metadata behoort een geïntegreerd deel van de dataset te zijn. De etiketten van de pindakaas zitten ook niet apart in een mapje achter de kassa van de supermarkt, ze zijn op de pot geplakt.

Om de toegankelijkheid te vergroten kan wel een catalogus worden gecompileerd, waarin alle metadata is samengebracht. Dat hoeft overigens geen geautomatiseerd systeem te zijn. Voordeel van een geautomatiseerde catalogus is wel dat deze door middel van een webservice op internet (intranet) gepubliceerd kan worden. Daarmee zijn de metagegevens eenvoudig toegankelijk.

Gebruikers kunnen zo snel vinden welke data waar beschikbaar is. De metadata geeft de gebruiker ook inzicht in de aard, actualiteit, kwaliteit en andere kenmerken van de data. Zo kan hij beoordelen of de gevonden dataset voldoet aan zijn vraag en hoe en onder welke condities hij erover kan beschikken. Aan het eind van de levenscyclus archiveert de bronhouder vervallen datasets en ook hier zorgt de metadata voor een adequate boekhouding.

Data en metadata zijn dus onlosmakelijk met elkaar verbonden. Met deze visie heeft ESRI metadata geïntegreerd in haar GIS producten. In ArcGIS beheert de gebruiker data, kaarten, modellen en metadata. Een Catalog Service is beschikbaar als onderdeel van de internet GIS server software omgeving, ArcIMS.

ArcCatalog: metadata in standaard GIS software

ESRI ondersteunt het creëren en bijhouden van metadata als standaardfunctionaliteit in ArcGIS desktop (in de applicatie ArcCatalog). Het beheer van metadata in ArcCatalog is volledig geïntegreerd met het databeheer. Sommige metadata velden zijn feitelijk fysieke kenmerken van het te beschrijven bestand. ArcCatalog genereert deze velden – 'properties' ofwel inherente metadata – automatisch. Dat bespaart veel werk en zorgt voor optimale metadata kwaliteit. Verandert de dataset, dan past ArcCatalog de properties aan. Data en metadata zijn dus altijd synchroon. Voorbeelden van properties zijn onder meer naam en formaat van de dataset, projectie, geografische extent, namen en technisch kenmerken van kolommen in de attribuuttabellen, aantallen objecten in een dataset alsmede diverse datumgegevens. De software gebruikt verschillende van deze properties, zoals de projectie, om de data op de juiste wijze te kunnen verwerken en presenteren.

De niet automatisch afleidbare metadata, de documentatie, wordt ingevuld met behulp van een editor. Voorbeelden van documentatie zijn definities, gebiedsaanduidingen en contactinformatie.

Metadata aanmaken blijft vervelend werk. Voor diverse velden heeft de editor in ArcCatalog daarom handige hulpmiddelen bij het invullen zoals defaults en keuzelijsten. Defaults en keuzelijsten zijn door de gebruiker (lees: metadata beheerder) samen te stellen. Leuker kunnen we het niet maken, wel makkelijker. De software geeft in de interface aan welke velden verplicht zijn maar dwingt het invullen niet af. Afdwingen leidt immers niet tot het beter invullen.

ArcCatalog slaat metadata op in het open XML formaat. Het belangrijkste voordeel van XML is dat er geen speciale software nodig is om XML te kunnen lezen. Een browser of Notepad volstaat al. Gegeven de aard en doelstelling van metadata zou het ook vreemd zijn als er speciale software nodig zou zijn om metadata te raadplegen.

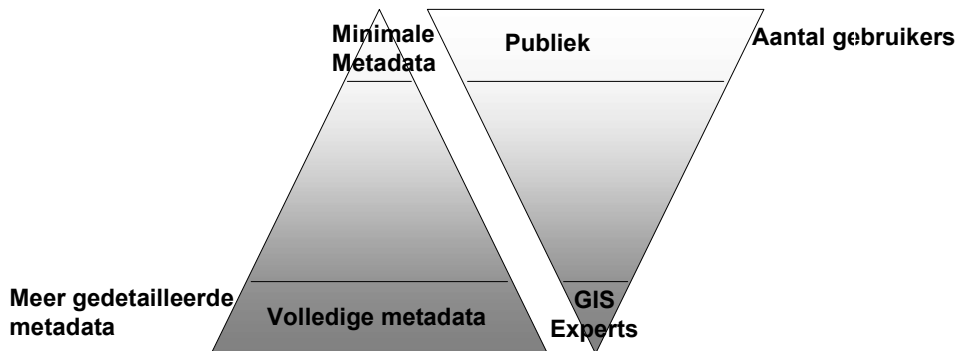
ArcCatalog kent verder het principe van stylesheets. In een stylesheet is gedefinieerd hoe een XML bestand wordt weergegeven. Een stylesheet is als het ware een filter op de XML. Door te schakelen tussen verschillende stylesheets zien de gebruikers naar behoefte meer of minder metadata. Of ze zien de metadata in een andere volgorde of opmaak. In ArcCatalog zijn uitgebreide zoekmogelijkheden aanwezig.

Voor wat betreft de inhoud van de metadata is ArcCatalog open. Het is niet gebaseerd op één bepaalde metadata content standaard maar het biedt de gebruiker een open raamwerk. Gangbare (internationale) metadata content standards als FGDC, CEN en ISO 19115 worden direct ondersteund maar men kan ook een eigen standaard implementeren. Er zijn diverse voorbeelden van nationale standards (Scandinavië, Australië) of domeinstandaards (geologie).

Verschillende behoefte van gebruikers

De behoefte die gebruikers hebben aan metadata verschilt. Grofweg is onderscheid aan te geven tussen:

- Verkennende metadata. Metadata op globaal niveau, vooral nodig voor het zoeken naar geschikte data.
- Gedetailleerde metadata (data dictionary). De uitgebreide metadata is nodig om gevonden data verder te beoordelen.



Metadata: verschillende behoefte.

Er is een omgekeerd verband tussen de omvang van de benodigde metadata en de aantallen gebruikers. Veel gebruikers hebben behoefte aan een beperkte set metagegevens, slechts enkele experts hebben behoefte aan alle details.

Het gaat er om dat gebruikers van metadata niet onnodig belast worden met technische details. Tegelijkertijd moeten zij wel adequate middelen ter beschikking hebben om data te kunnen vinden. Een goede metadata oplossing voorziet in deze verschillen in behoeften. In ArcCatalog is het mechanisme van de stylesheets hiervoor heel geschikt gebleken.

Op een portaal zoekt een gebruiker in principe in een beperkt aantal velden (verkennende metadata). Van de gevonden datasets die aan de opgegeven globale zoekcriteria voldoen worden vervolgens detailgegevens opgevraagd. Overigens kan ook hier het stylesheetprincipe worden toegepast.

Webservices

Webservices hebben recent grote veranderingen in de informatietechnologie teweeggebracht. Interessanter dan de techniek als zodanig is wat er feitelijk mogelijk wordt door de toepassing ervan. Met webservices kunnen op uiterst efficiënte wijze gedistribueerde informatiesystemen worden opgezet. Het is een zeer geschikte techniek voor het uitwisselen van informatie tussen systemen en tussen organisaties.

Data blijven beheerd bij de bron en gebruikers tappen via internet de brondata af. De voordelen zijn evident. In de eerste plaats is data direct van de bron altijd actueel en van gegarandeerde kwaliteit. Alle gebruikers van de betreffende bron werken met dezelfde versie van de gegevens. Daarnaast levert het een aanzienlijke besparing in beheerkosten.

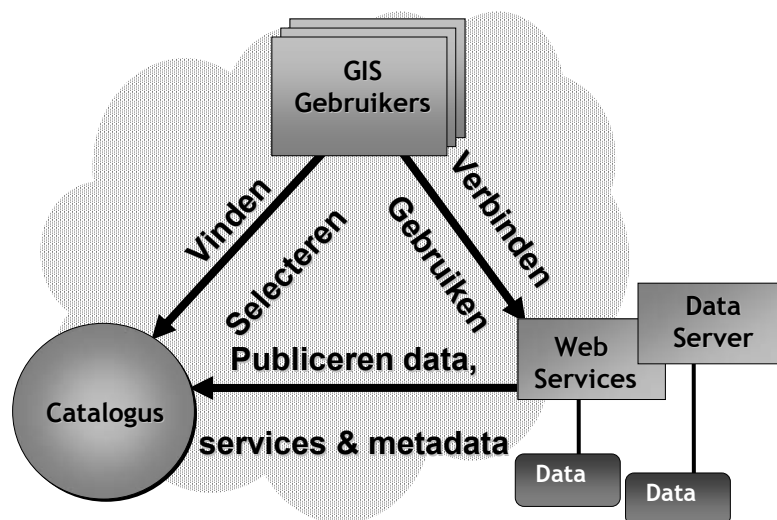
De potenties van webservices voor GIS zijn groot. Er zijn al diverse succesvolle voorbeelden. Leveranciers van grote algemeen gebruikte datasets (topografie, luchtfoto's, stratenbestanden) publiceren hun informatie in de vorm van webservices. Ontwikkelaars van internetsites bouwen op basis van zo'n service GIS functionaliteit of kaartjes in in hun GIS webapplicaties.

Een ander voorbeeld van een (potentiële) toepassing is de uitwisseling van planinformatie tussen gemeente en provincie. De gemeente publiceert in een zo vroeg mogelijk stadium planinformatie van een nieuw bestemmingsplan. Aan de andere kant publiceert de provincie 'toetslagen'. Doordat beide partners in het planvormingsproces al tijdens de ontwerpfase relevante informatie met elkaar delen verloopt het proces efficiënter en is het resultaat een beter afgestemd plan.

De laatste versies van desktop GIS pakketten als ArcView zijn in staat om verbindingen te leggen met webservices. Dit betekent dat gebruikers van deze pakketten data uit externe bronnen integreren in hun eigen werkomgeving en met hun eigen data. Er zijn dus geen specifiek ontwikkelde client-applicaties meer nodig om met webservices te kunnen werken. GIS is gedistribueerd.

Dat metadata in een dergelijke omgeving van vitaal belang is mag evident zijn. Wat heb je aan een service als je niet weet wat de inhoud is? Overigens, de laatste ontwikkeling is dat services 'self describing' zijn.

In een gedistribueerde GIS-omgeving brengt metadata aanbieders en afnemers van data(services) bij elkaar. De aanbieder beschrijft zijn bron met metadata en publiceert het aanbod in een catalogus. Gebruikers zoeken in de catalogus naar data die aansluit bij hun vraagstelling. Na een eerste inspectie van potentiële bronnen – uiteraard door de metadata goed te bestuderen – selecteert de gebruiker databronnen van zijn keuze en maakt verbinding met de service om de data te gaan gebruiken.



Gedistribueerd GIS.

Dit gedistribueerde model is onderliggend aan een Geo Informatie Infrastructuur (of: Spatial Data Infrastructure – SDI). Boeiend is dat het model op diverse schaalniveaus toepasbaar is.

Binnen een organisatie spreekt men van een Concern SDI, op nationaal niveau van de Nationale Geo Informatie Infrastructuur en wereldwijd van de GSDI (Global Spatial Data Infrastructure).

Catalogus

Binnen een gedistribueerde GIS omgeving zijn services beschreven met metadata; metadata is een onderdeel van de informatiebron. Een niet geringe uitdaging is het doorzoekbaar maken van de gedistribueerde documentatie. Hier speelt de catalogus (ook een service) een essentiële rol.

Voor het opzetten van een catalogus bestaan in principe twee concepten. Het eerste model gaat uit van een centrale ingang naar de gedistribueerde bronnen. Vanaf de centrale ingang stuurt men een zoekactie uit over de verschillende bronnen.

De andere benadering noemt men 'harvesting'. Daarbij worden de metagegevens uit de verschillende bronsystemen samengebracht in één centrale catalogus.

Beiden benaderingen hebben voor en nadelen. Het gedistribueerde model heeft als voordeel dat de metadata maar op één plaats beheerd wordt maar is technisch complex. Op alle bronsystemen dient een catalogus service operationeel te zijn. Elke zoekactie komt op elke geregistreerde server. De overall performance van een zoekactie wordt bepaald door de langzaamste server. Er is geen controle over de kwaliteit van de metadata. De ervaring leert verder dat ondanks het gebruik van metadata standaards er toch inhoudelijke verschillen in de metadata zijn. Datumnotaties kunnen verschillen, domeinen van metadatavelden als onderwerp of thema zijn vaak per organisatie anders etc. Dat maakt het lastig, zo niet onmogelijk om gedistribueerd te zoeken.

Een belangrijk voordeel van 'harvesting' is dat daarmee een mogelijkheid ontstaat om bij het opladen naar de centrale catalogus de hiervoor genoemde verschillen op te lossen. De metadata content is heel nauwgezet te standaardiseren. Bij het opladen van de metadata van de bronnen naar de centrale catalogus vindt controle en validatie plaats. Verder garandeert een schaalbare technische infrastructuur onder de centrale catalogus stabiliteit en een goede performance.

Tenslotte

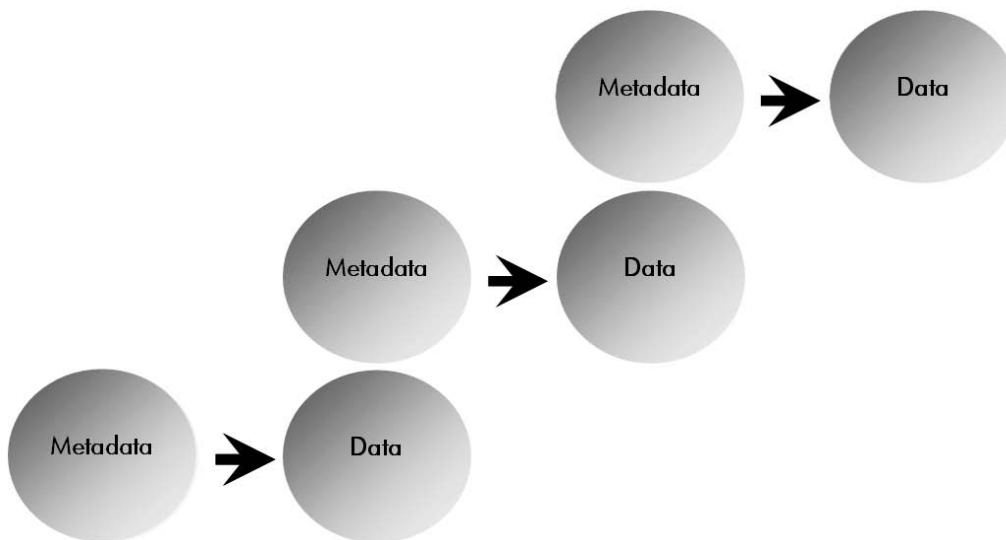
De technologie van webservices is veelbelovend. Webservices geven – eventueel via een portaal – toegang tot een veelheid aan bronnen. Er ligt een wereld aan informatie klaar om gebruikt te worden. Maar is dit dan werkelijk de Haarlemmerolie voor alle vraagstukken? Technisch is het geen probleem meer om informatie van uiteenlopende oorsprong samen te brengen. Wat resteert, is de afstemming van de inhoud van verschillende bronnen. Op dat traject is echter nog een lange weg te gaan. Zolang inhoudelijke standaardisatie nog ver weg is, zal metadata moeten zorgen voor een adequate beschrijving van een bron. Daarmee zijn afnemers van de informatie in staat te beoordelen of een service (of bestand) aansluit bij hun eisen.

De Universal Data Store en de rol van meta-informatie

Ir. Remco J.J.H. van Eijndt
URBIDATA bv
remco.van.eijndt@urbidata.nl

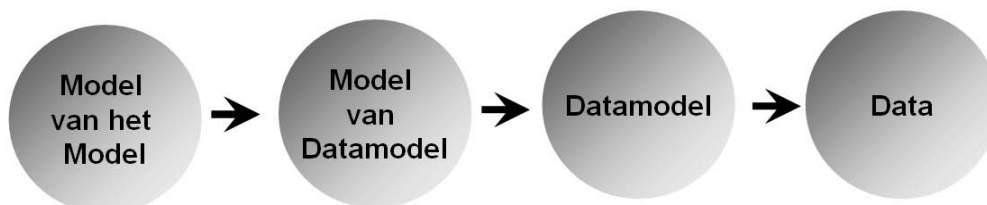
Wat zijn metadata?

Ten behoeve van elke zinvolle vorm van communicatie is het noodzakelijk om een gemeenschappelijke basis te hebben. Bij communicerende software zal deze ondubbelzinnig vastgelegd moeten zijn. Dit in tegenstelling tot het dagelijkse taalgebruik, waarbij we zeer uiteenlopende informatie verkrijgen die ons helpt bij het bepalen van de semantiek. Voorbeelden van deze uiteenlopende informatie zijn plaats, tijdstip en het weer. Software werkt met gegevens met daarbij, als het goed is, de complete en correcte metadata.



Figuur 1. Metadata is een relatief begrip.

Metadata leggen de structuur en semantiek vast van een dataset. Deze metadata kunnen we vervolgens uitwisselen om zodoende een gemeenschappelijke basis te realiseren.. Verder zal het registreren van metadata de eigenaar van de data dwingen na te denken over de exacte betekenis om te komen tot een ondubbelzinnige, toegankelijke dataverzameling. Het is belangrijk om in te zien dat metadata een relatief begrip is, zie figuur 1. Over elke gegevensset kunnen gegevens worden vastgelegd, dus ook over metagegevens, metametagegevens, et cetera.



Figuur 2. Metadata op verschillende niveaus.

Als voorbeeld staat in figuur 2 de gegevensvastlegging op vier niveau's in relatie tot de Universal Data Store (UDS):

- *Data*. De brongegevens die worden opgeslagen in de Operational Data Store (ODS) en het Data Warehouse (DWH) van de UDS.
- *Datamodel*. De datamodellen van de bronnen en van de ODS en het DWH zijn metadata van de UDS. Deze gegevens worden opgeslagen in de Metadatabase.
- *Model van datamodel*. De metadatabase heeft zelf uiteraard een datamodel. Dit datamodel is zo ontwikkeld dat het de opslag van beschrijvingen van onder andere brondatamodellen mogelijk maakt. Het datamodel van de metadatabase is omschreven in UML-notatie.
- *Model van het model*. Het model van een UML-diagram is de definitie van UML zelf.

Uiteraard is dit overzicht niet compleet. De definitie van UML kan wederom in UML zelf of in een willekeurige andere notatievorm worden gegeven. Hieruit blijkt duidelijk dat metadata een relatief begrip is. In feite kan gesteld worden dat gegevens omschreven moeten worden middels een context om er betekenis aan te geven. Deze context dient vervolgens wederom omschreven te worden. In de praktijk zal de contextvastlegging zich uiteraard tot een bepaald niveau beperken.

Metadata bij transformaties

Om het belang van metadata te onderstrepen bij het transformeren van gegevens, zal eerst de theoretische kant van transformaties bekeken worden. Bij het transformeren van (geografische) data kan onderscheid worden gemaakt naar enerzijds 'structural transformation' en anderzijds 'semantic translation'.

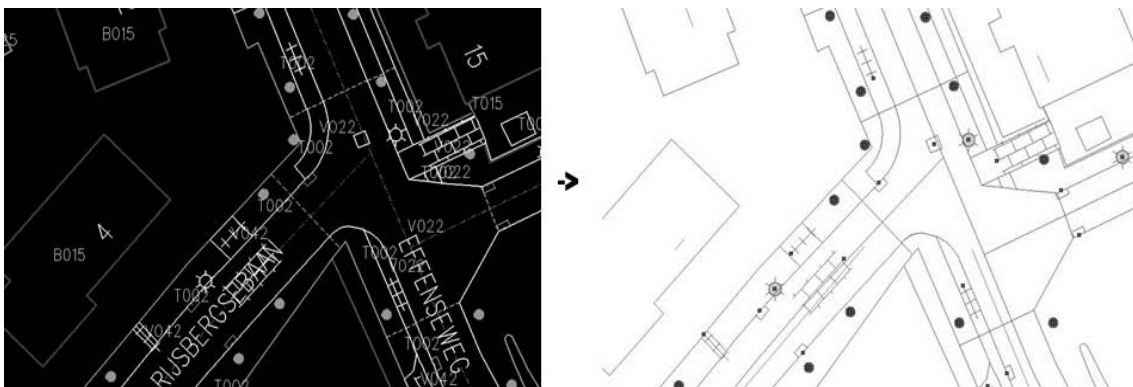
Structural transformation

Meestal wordt bij het beschrijven van een uitwisseling van geografische data een structural transformation beschreven. Bij deze vorm van transformatie wordt vastgelegd welke verplaatsing en (structuur-) verandering de data ondergaan. De geografische data worden aan het ene datamodel onttrokken en zo geconverteerd dat ze passen in het andere datamodel. Er kunnen drie verschillende vormen onderscheiden worden:

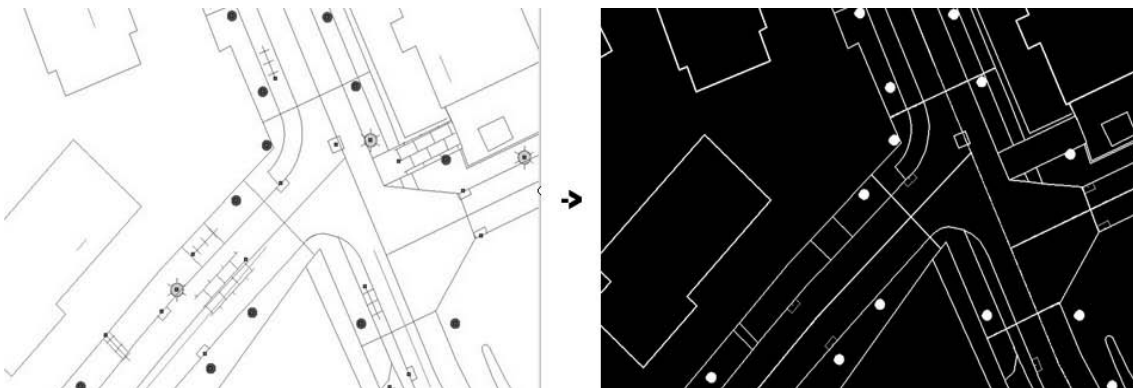
- *Formaat converters*. De eenvoudigste benadering is de directe formaatconversie. Hierbij wordt rechtstreeks geconverteerd van het datamodel en -formaat van het ene systeem naar het andere systeem. Deze manier van transformeren is meestal zeer efficiënt omdat aangenomen mag worden dat er uitgebreide kennis is van het bron- en doelsysteem. Vaak kunnen data wel van het bron- naar het doelsysteem worden getransformeerd, maar niet vice versa. Dit is het geval als er sprake is van verlies aan informatie doordat data in het doelsysteem minder uitgebreid worden opgeslagen dan in het bronsysteem. Een nadeel van deze strategie is dat er voor n verschillende systemen, n^2 converters nodig zijn.
- *Standaardformaten*. Om het n^2 -probleem op te lossen, is er een aantal standaarden ontworpen. Sommige standaarden zijn door systeemontwikkelaars ontworpen voor het eenvoudig uitwisselen van data ten behoeve van hun eigen systemen.
- *Uitwisselingsformaten*. In tegenstelling tot productgeoriënteerde standaardformaten zijn er uitwisselingsformaten ontwikkeld. Deze uitwisselingsformaten zijn bedoeld als heldere en duidelijke formaten die data kunnen verwerken uit verschillende datamodellen en -formaten. Uitwisselingsformaten zijn meestal ontwikkeld door een samenwerkingsverband van verschillende specialisten op één bepaald gebied. Hierdoor ontstaat echter wel weer een nadruk

op één bepaald soort informatie. De uitwisselingsformaten zijn tevens vaak gebaseerd op de grootste gemene deler van veel standaardformaten.

Dat een structuurtransformatie zich slechts richt op het 'vertalen' van de syntax van de data geeft gelijk de beperking weer van een dergelijke conversie. Het betreft een zuiver geometrische omzetting van gegevens waarbij vaak informatie verloren gaat ten aanzien van de cartografische aspecten van een dataset. De kaartopmaak geeft betekenis aan de inhoud van een dataset door middel van kleurgebruik, lagenindeling, lijnstijlen et cetera. De structuur van de cartografische opmaak verschilt per formaat zeer sterk. Hierdoor gaan deze gegevens bij een structuurtransformatie van het ene naar het andere formaat vaak verloren. Door deze incompatibiliteit kunnen formaten deze gegevens niet uitwisselen en beperken zich tot het omzetten van punten, lijnen en vlakken.



Figuur 3. Transformatie van structuur A naar structuur B.



Figuur 4. Transformatie van structuur B naar structuur A.

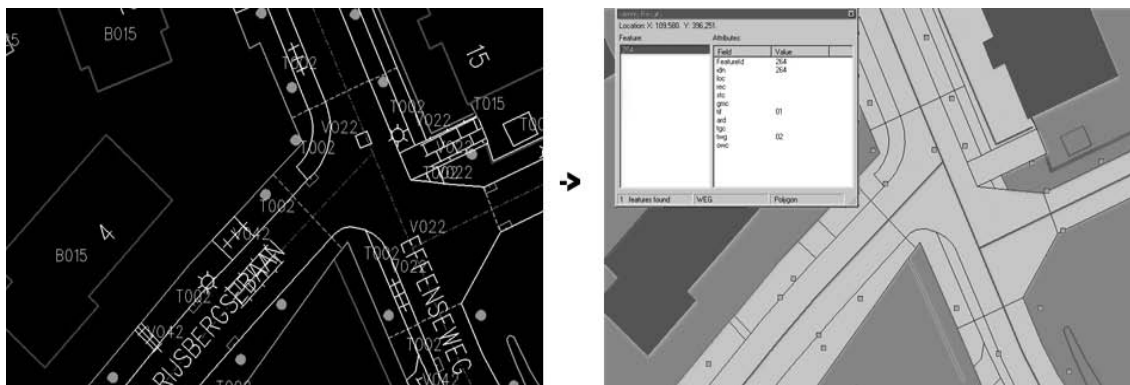
In figuur 3 en 4 is het resultaat te zien van een structuurtransformatie tussen de structuur van formaat A en B. De data zijn van A naar B en vervolgens weer naar A getransformeerd. Uit de figuren komt duidelijk naar voren dat slechts het gemeenschappelijk model van beide structuren, te weten het punten-, lijnen- en vlakkenbeeld, overblijft. De ontwikkelde conversies voor de formaten kunnen slechts communiceren op basis van wat men een Jip-en-Janneke taal zou kunnen noemen.

Semantic translation

In tegenstelling tot de uitwisseling van de structuur van de geografische data is de uitwisseling van de betekenis van de gecodeerde geografische data vaak slecht gedefinieerd. De betekenis van geografische data komt tot uiting in de thematische beschrijving van elk ruimtelijk object.

Als beide datamodellen, die betrokken zijn bij een transformatie, dezelfde termen, definities en classificaties gebruiken, zal de semantische vertaling voor honderd procent slagen. In deze situatie houdt de transformatie slechts in dat de databasevelden in een overeenkomstig veld, met dezelfde definitie, gezet hoeven te worden. Echter in de praktijk zijn er grote verschillen in definities en termen die gebruikt worden in verschillende datamodellen voor dezelfde reële objecten. Ook worden verschillende reële objecten in het ene datamodel verschillend benoemd, terwijl ze in het andere dezelfde definitie krijgen. Als blijkt dat definities wel overeenkomen, dan is het nog vaak zo dat een object in het ene datamodel gedetailleerder wordt omschreven dan in het andere.

Semantische vertaling zal dus plaatsvinden door middel van het converteren van de thematische beschrijving van een ruimtelijk object van de ene classificatie naar de andere. Met behulp van een classificatie worden ruimtelijke objecten gecategoriseerd volgens de rol die de objecten hebben binnen een informatiesysteem. Objecten met gelijke rol vallen in een zelfde categorie. De rol van een object wordt bepaald door zijn thematische beschrijving. Een klasse wordt hier dus in verband gebracht met de beschrijvende structuur van een object. Objecten die binnen een zelfde klasse vallen, bevatten dezelfde soort informatie.



Figuur 5. Transformatie van A naar B inclusief volledige semantiek omzetting.

De semantiek van objecten is vooral bij CAD-formaten vaak terug te vinden in de cartografische opmaak van de kaart. In GIS-omgevingen waarbij de gegevens in een meer databaseachtige omgeving wordt opgeslagen is deze semantiek terug te vinden door de thematische attributen die aan objecten gekoppeld worden.

Bij de vertaling van een CAD-bestand naar een GIS-omgeving dient deze opmaak op een juiste manier vertaald te worden naar attribuut informatie. In figuur 5 is een voorbeeld te zien van een 'slimme' omzetting van een cartografisch rijke kaart naar een GIS-bestand. Van het lijnenbeeld is een vlakkenkaart gecreëerd waarbij tevens de lijnstijlgegevens zijn vertaald naar attribuut informatie. Met behulp van een uitgebreide verzameling 'transformatieregels' kan op deze manier een objectgerichte kaart gemaakt worden. In het voorbeeld is in het bronbestand gebruik ge-

maakt van een LKI-codering voor de geometrie. Het objectgerichte model van het doel is gebaseerd op de classificatie volgens het Informatie Model Ruimtelijke Ordening (IMRO).

Views en transformatieregels

Semantic translation maakt het dus mogelijk om uitgebreid data te herstructureren. Er kunnen in feite meerdere 'views' gecreëerd worden op dezelfde data. Een view is een bepaalde kijk op gegevens die niet op deze manier fysiek is gemodelleerd maar wel kan worden afgeleid. Zo kan aan een zelfde gegevensstructuur inhoud worden gegeven ten behoeve van verschillend gebruik. Er kan als het ware betekenis worden toegevoegd aan de structuur zodat de data in een andere context geplaatst kan worden. Dit is bijzonder handig indien men de data wil ontsluiten richting andere gegevensbronnen. Voordat gegevens in een andere bron kunnen worden gebruikt, dienen ze meestal getransformeerd te worden. Bij het transformeren van data is het noodzakelijk om regels vast te leggen over hoe deze transformatie precies dient plaats te vinden. Een eerste stap hierbij is het bepalen van een gemeenschappelijke taal zodat op basis hiervan gegevens uit een bron kunnen worden vertaald naar een doel. Met het creëren van views op de data, of datamodellen, kan deze gemeenschappelijke taal nu juist worden vastgelegd.

Metadata in de UDS

Binnen de Universal Data Store (UDS) worden verschillende soorten metadata geregistreerd met behulp van de UDS MetadataManager:

- *Catalogus*. Bevat een volledige omschrijving van datasets in de vorm van gegevenswoordenboeken. Een gegevenswoordenboek is méér dan alleen een datamodel. Het bevat tevens informatie over alle definities, betekenis van aanwezige relaties, invulling van domeinen et cetera.
- De UDS MetadataManager maakt het mogelijk om te zoeken in de catalogus op basis van willekeurige termen en eigenschappen en biedt tevens ondersteuning bij het stellen van complexe vragen aan de data (rapportages).
- *Transformaties*. Een transformatie omschrijft een afbeelding tussen twee verschillende datamodellen die in de catalogus worden vastgelegd. Een transformatie bestaat in feite uit een groot aantal afzonderlijke transformatieregels die de precieze relatie aangeven tussen een verzameling objecten uit een bron en een verzameling objecten uit het doel van de transformatie. Deze relatie kan vorm worden gegeven door middel van een functieomschrijving in een scripttaal zoals Javascript of VBscript.
- *Processen*. Een proces geeft een beschrijving van de uitvoering van transformaties. Met behulp van de ingevoerde transformatieregels kunnen processen worden beschreven die semantic translation, zoals in het vorige hoofdstuk beschreven staat, mogelijk maken.